

Beyond the cloning debate

Arguments over human embryonic stem cells and cloning have loomed large over the choice of a new director of the National Institutes of Health. But the expected nominee for the position will need to prove himself on other fronts.

Within the next month or two, the US Senate will vote on a bill that would ban research into therapeutic cloning — the creation of cloned human embryos to extract embryonic stem (ES) cells matched to individual patients. The result is too close to call, and lobbying for and against has reached fever pitch.

In this febrile atmosphere, unpublished scientific claims are assuming a weight usually afforded to peer-reviewed results (see *Nature* **415**, 722; 2002). Last week, for instance, *The Wall Street Journal* discussed unpublished claims by Chinese researchers to have extracted ES cells from cloned embryos created by fusing human cells with rabbit eggs. Supporters of therapeutic cloning argued that other nations stand to reap the economic benefits if the United States turns its back on the field; opponents seized on the 'yuk' factor of combining human and animal material.

Even basic terminology is being reinvented. While the anti-abortion lobby characterizes therapeutic cloning as 'clone and kill' medicine, the presidents of the National Academy of Sciences and the Institute of Medicine argue that 'therapeutic cloning' should be abandoned in favour of the phrase 'somatic cell nuclear transfer' (*Science* **295**, 1237; 2002). The latter phrase is conceptually accurate, and in avoiding the word 'cloning' makes a clear distinction between reproductive cloning and the therapeutic procedure. But it hardly trips off the layperson's tongue, and may create the impression that scientists want to hide the fact that the procedure involves creating an embryo — albeit one consisting merely of a hollow ball of cells.

This is the background against which President George W. Bush is expected to announce this week that Elias Zerhouni, executive vice-dean of the Johns Hopkins University School of Medicine in Baltimore, Maryland, is his nominee for the post of director of the National Institutes of Health (NIH). Zerhouni was reportedly selected ahead of Tony Fauci, the widely tipped director of the National Institute of Allergy and Infectious Diseases (NIAID), because he was willing to endorse Bush's support for a ban on therapeutic cloning. Fauci, it is alleged, was insufficiently 'pro-life' to be trusted with the top NIH post.

Wisely, Zerhouni is keeping his own counsel. But when the time for his confirmation hearings in the Senate comes round, it is to be hoped that his position on ES-cell research and cloning does not dominate the discussion to the exclusion of all else.

Once US federal policy on cloning has been decided, Zerhouni's personal position will be moot. But the other issues facing the NIH will remain. Departed senior staff will still need to be replaced; debates about priorities in genome research will still need to be resolved; and the new director must develop a fruitful working relationship with Fauci so that the vast injection of funds into the NIAID for research into countering bioterrorism is wisely spent.

The senators who will debate Zerhouni's appointment should put his position on cloning to one side, and concentrate on examining how his experience at Johns Hopkins qualifies him to deal with the rest of the NIH's agenda. ■

Helping hands for Arab science

A grassroots initiative to boost research in the Middle East deserves support — in particular from Arab Americans.

Later this month in Sharjah, in the United Arab Emirates, some 500 scientists from across the Arab world will hold what they hope will be a historic gathering. They will set research priorities for the Arab Science and Technology Foundation (ASTF), which aims to raise up to US\$150 million over five years to award peer-reviewed grants to the region's scientists (see pages 120–122).

The foundation's leaders face an enormous task. Raising large sums of money is always hard, but raising the sum the ASTF has in mind, while retaining independence, will be doubly difficult. Not only have most Arab governments long neglected science, but many also distrust organizations over which they do not exert direct control.

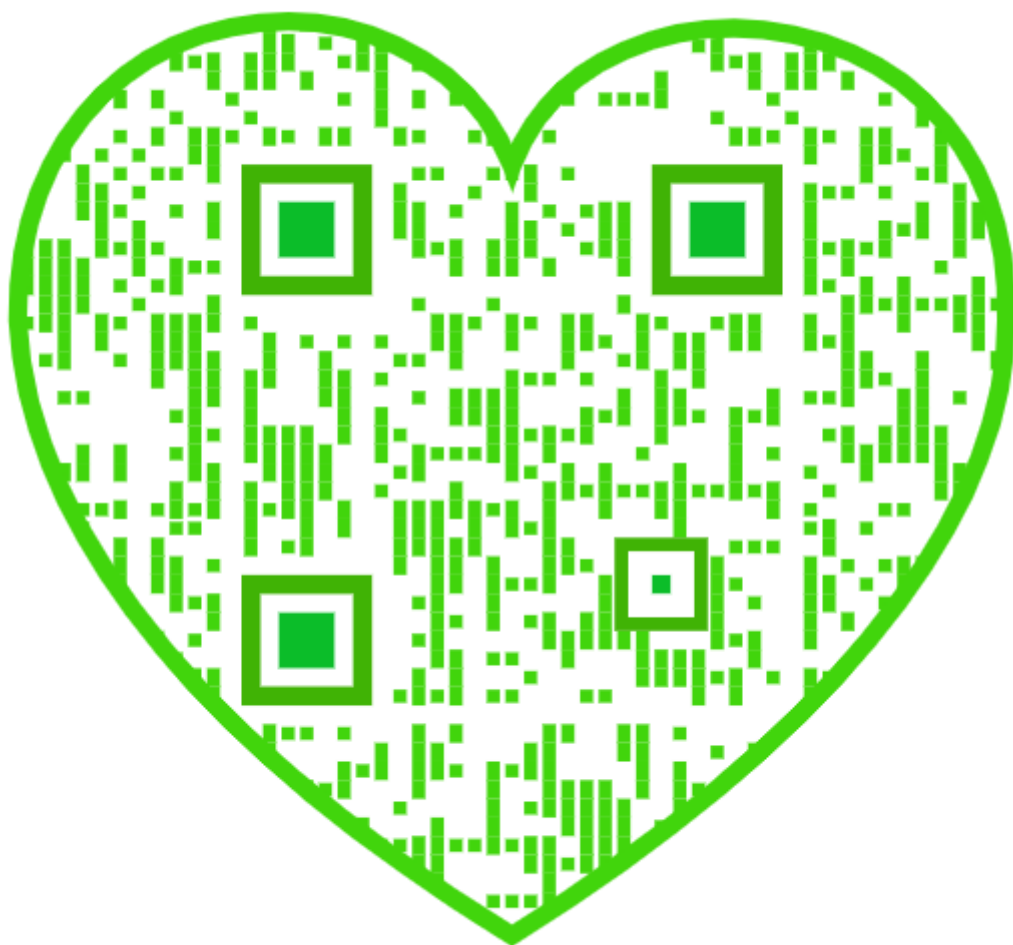
The initial signs are encouraging. The ASTF has built itself from the bottom up, gathering together working scientists interested in strengthening research capacity across the region. It is committed to building links between Arab and Western labs; it has already raised some \$6 million from the ruler of Sharjah, who has promised not to interfere with the running of the foundation; and it has secured an energetic cheerleader in Mohammed Aref, formerly science editor of the London-based *Al Hayat* newspaper, who several weeks ago visited *Nature*'s London office and made a passionate case for the ASTF.

Western research agencies appear keen to build links with Arab

scientists, and hopefully will work with the ASTF. But cooperation at the level of government agencies will remain vulnerable to the region's troubled politics — some of those planning to attend the Sharjah meeting, for instance, are coming from Iraq, a country that may soon be at war with the United States.

For this reason, the ASTF should also look to individual donations. Here, those of Arab descent living in the United States may prove useful. Particularly in the industrial Midwest, Arab Americans are an emerging economic and political force, courted by President George W. Bush in his 2001 election campaign. It is also worth noting that the scientist expected to be nominated by Bush for the post of director of the National Institutes of Health (see above) was born and educated in Algeria.

It may be a provocative example, but the Weizmann Institute of Science in Rehovot, Israel, hints at the potential. Around 17% of its annual budget of about \$180 million comes from donations, more than half being raised by the American Committee for the Weizmann Institute of Science — mostly from Jewish Americans. A tradition of scientific philanthropy isn't built overnight, but the ASTF has much to gain from convincing wealthy Arab Americans to rival their Jewish counterparts in their support of research. ■





Hubble trouble
Astronomers seek
telescopic extension
past 2010
p112



Two into one
Manchester's
universities may
join forces
p114



Insect nation
Ants identified as
species database is
launched
p115



Poached eggs
Caviar back on the
menu for traders at
the Caspian Sea
p116

Geneticists get steamed up over public access to rice genome

Declan Butler

Twenty top genome researchers have written to the editorial advisers of *Science* protesting at the way the journal occasionally publishes genome maps without requiring the authors to place the supporting sequence data in public databases.

The letter is signed by such luminaries as Bob Waterston, head of genetics at Washington University in St Louis, Nobel laureate Aaron Klug of the MRC Laboratory of Molecular Biology in Cambridge, UK, and Michael Ashburner, former head of the European Bioinformatics Institute at Hinxton near Cambridge. In it they argue that new genome sequences should be made available in public-domain databases in line with what they term "accepted norms of the field".

"There are strong rumours in the field that *Science* is considering allowing the publication of papers from commercial companies on the rice and mouse genomes, without demanding the submission of the data in GenBank as a condition," their letter says.



Boiling point: disputes about gene data have spilt over to the planned publication of a rice genome.

Several sources confirm that *Science* intends to publish a paper by the Swiss-based agricultural biotechnology company Syn-

genta on its draft of the rice genome. The supporting sequence data will not be deposited in GenBank, the sources say, but will be available free to academic researchers from Syngenta's website, subject to certain restrictions.

Science drew criticism last year when it agreed to publish the draft human genome assembled by Celera Genomics of Rockville, Maryland, despite the company's restrictions on access to the sequence data.

Donald Kennedy, *Science*'s editor-in-chief, declines to comment on the pending paper. "*Science* is committed to full public access," he says. "But we will consider rare exceptions if the public benefits of removing valuable data and results from trade-secret status clearly exceed the costs to the scientific community of the precedent the exception might create. This was true for the human genome sequence, and for the most important agricultural commodity in the Third World, the case is surely even stronger."

According to several researchers, *Science* also plans to publish a draft sequence of *Oryza sativa* L. ssp. *indica*, the major crop rice cultivar in China, alongside the Syngenta genome. This second rice genome was completed recently by a team led by Huanming Yang, director of the Beijing Genomics Institute, and the supporting sequence

Societies query student-visa review

Geoff Brumfiel, Washington

Scientific societies say that they have been locked out of a government task force that is drafting restrictions on the flow of international students into the United States in the wake of last autumn's terrorist attacks.

The task force on student visas is expected to release draft proposals within the next month for the entry of both undergraduate and graduate students. It is made up of officials from the White House Office of Homeland Security and the Office of Science and Technology Policy, together with representatives from the state and justice departments, which oversee immigration.

Irving Lerch, head of international affairs at the American Physical Society, says that universities and societies are not involved in the debate. "When you have such

important discussions going on within the government, without the participation of those who know what the problems are, you have a prescription for disaster," he says.

Lerch fears that the final criteria will curtail recruitment from 'sensitive' countries such as India and China, which supply the United States with large numbers of graduate students.

The closed nature of the process is "of enormous concern," says Vic Johnson, director of public policy at the Association for International Educators, a Washington-based group that promotes student exchanges. "Let's not go overboard and cut off scientific exchange," he says.

But the White House says that the panel's findings will be discussed with universities before they are implemented. "As soon as we have something we can share, we plan to share it," a White House official says. ■

data have been deposited in GenBank.

A draft sequence of the rice genome by the agricultural biotechnology company Monsanto, based in St Louis, Missouri, and one by Celera of the mouse genome, are also under preparation, but have not yet been scheduled for publication in any journal.

Syngenta currently makes its data available to a handful of academic groups through special agreements. The publication of Syngenta's rice genome in *Science* might result in changes to the company's policy, giving more researchers access to the sequence data. But, as the letter demonstrates, researchers remain deeply divided over the terms of such access. "This goes to the heart of what science is all about, the free exchange of ideas, data and reagents," says Bruce Stillman, director of the Cold Spring Harbor Laboratory in New York state. *Science* should not compromise on making the data freely available, he says.

But Ron Cantrell, director of the International Rice Research Institute in the Philippines, is more supportive of *Science*'s decision to publish. "You have to ask the question 'is it better not to have any access at all?'," he says, adding that, in his experience, Syngenta and Monsanto have "been very forthcoming" in collaborations with the public sector.

Chris Novak, a spokesman for Syngenta, says that the company hopes to work with the publicly funded International Rice Genome Sequencing Project (IRGSP). The project intends to produce a 'finished' high-quality sequence, as opposed to the drafts, containing many gaps, that are about to be published.

Researchers point out that *Science*'s agreement with Syngenta is not entirely analogous to the one it reached last year with Celera on the human genome. Celera contributed no data to the public Human Genome Project, instead relying on data from the public project to complete its own sequence. In contrast, Syngenta has already contributed significant mapping data to the IRGSP, through a collaboration with Clemson University in South Carolina.

But Syngenta has so far refused to share its raw sequence data with all of the public group — unlike its rival Monsanto, whose contributions of sequence data are credited with strongly accelerating the public project.

In January, however, Syngenta began talks with the IRGSP and, according to one IRGSP official, has agreed in principle to match the Monsanto agreement. If it does, "all the Syngenta and Monsanto data will be in the public domain by the end of the year," says the official. The likelihood of this happening might be a factor in persuading *Science* to accept restrictions on the rice data for the time being, observers suggest.

NASA urged to play waiting game on Hubble's retirement

Tony Reichhardt, Washington

Even as NASA completes a successful service mission on the Hubble Space Telescope, a debate is simmering in the agency about how long to keep the instrument in action.

NASA's current plan is to return the telescope to Earth in 2010 and put it in the Smithsonian National Air and Space Museum in Washington. But some Hubble-project scientists and engineers at NASA's Goddard Space Flight Center in Maryland have been arguing — albeit with little success so far — that the telescope should continue its observations.

The dilemma of when to pull the plug on a successful mission is a familiar one at NASA. But the case of Hubble will be especially difficult. Not only is it the most powerful telescope in history, but its observing grants are now established as a mainstay of support for US astronomers. And after last week's service mission — the fourth since its launch in 1990, and one which brought an infrared camera back to life and installed the powerful Advanced Camera for Surveys — the telescope has never been more capable.

But NASA needs to retire Hubble to pay for its successor, the Next Generation Space Telescope (NGST). The agency has already delayed Hubble's retirement from 2005 to 2010, and is reluctant to do so again to appease those whom former NASA administrator Dan Goldin once derided as "Hubble huggers". So the current plan is for astronauts to install another camera and spectrograph during a fifth and final servicing mission in 2004, and to bring the telescope home six years later.

But project engineers at Goddard say that this will not be straightforward. Dismantling Hubble and returning it to Earth is more challenging than previously thought. The mission would require up to five spacewalks, and the telescope's weight would strain the shuttle's maximum landing limit.

NASA had hoped to launch the NGST



Despite the telescope's renewed capacity, NASA is reluctant to spend money on Hubble as it ages.

before Hubble retires, but the launch date has already moved from 2007 to 2009, and could easily slip further. According to Goddard engineers, Hubble is unlikely to survive six years after 2004 without maintenance.

So the Goddard team has proposed that another service mission be added in 2007. As well as making any necessary repairs, shuttle astronauts would attach a propulsion module that could allow the telescope to be 'de-orbited' safely to burn up in the atmosphere.

The Air and Space Museum would lose an exhibit, but scientists would gain more Hubble viewing time. Hubble senior project scientist Dave Leckrone of Goddard, who advocates extending the telescope's life, admits it will be an uphill struggle to convince NASA. "The official position is that there will be no more servicing of Hubble after 2004," he says.

A committee representing Hubble users advised NASA last October to consider the extra servicing mission. Committee chairman George Miley of Leiden University in the Netherlands agrees that delaying the NGST would be "bad for astronomy", but adds that "the effectiveness of an extra Hubble refurbishment mission should be seen in a wider context and considered within the framework of the NASA programme as a whole".

But in a letter to Anne Kinney, director of NASA's 'Origins' programme, an advisory committee to that programme argued that "minimizing additional expenditures on [Hubble] is crucial to keep the development of NGST on track".



Radiologist in the picture for top job at NIH

Erika Check, Washington

Elias Zerhouni, a radiologist and senior administrator at the Johns Hopkins University School of Medicine in Baltimore, Maryland, is set to be nominated by President George Bush as director of the National Institutes of Health (NIH).

If his nomination goes ahead this week, as government and university officials anticipate, and is confirmed by the Senate, Zerhouni will fill a position that has been vacant since December 1999, when Harold Varmus departed to become president of the Memorial Sloan-Kettering Cancer Center in New York.

Varmus is a Nobel prizewinning molecular biologist, and Zerhouni has never worked in basic research. But supporters say that the 50-year-old Algerian would bring his own mix of relevant skills to the administration of the world's largest research agency.

Zerhouni, a naturalized American citizen who is married with three children and lives in Pasadena, Maryland, earned his medical degree from the University of Algiers before coming to Hopkins as a resident in 1975. He helped to pioneer the use of new techniques in clinical radiology at the medical school



Elias Zerhouni is seen by supporters as an innovator with an administrative track record.

and worked in several administrative positions there. He also founded a company that sells devices for use in high-resolution magnetic resonance imaging of the body.

Colleagues say that Zerhouni has a warm, engaging personality, adding that he is seen as an innovator who devised techniques such as tagged magnetic resonance imaging, a non-invasive way to monitor heart contractions.

John Griffin, chair of neurology at the medical school, says that Zerhouni was an administrator with "vision" who also made decisions fairly, based on the data and faculty input. Scientists at Hopkins say that Zerhouni, as executive vice-dean of the medical school, has paid close attention to their needs.

Zerhouni's experience at Hopkins, his supporters argue, would stand him in good stead for the administrative challenges of running the NIH. "The NIH is just a bigger version of Johns Hopkins," says Hopkins neuroscientist Jeffrey Rothstein. "Hopkins and the NIH are each institutions that believe strongly in 'bench to bedside'—which means you do the basic research and are surrounded by colleagues who do the clinical research."

There is little public record of Zerhouni's political views, but he was a consultant to Ronald Reagan's White House from 1985 to 1988 where, some observers speculate, he may have made the connections that could lead to his nomination.

More recently, Zerhouni has been involved in the planning of a new \$80-million Institute for Cell Engineering at Hopkins, which will conduct interdisciplinary research on stem cells when it opens next year. But Griffin, who has also been involved in planning the institute, cautions that Zerhouni's involvement there doesn't signal an endorsement of experiments on human embryonic stem cells.

However, another researcher familiar with the institute, who didn't want to be identified, says that there is a "bit of a disconnect" between Zerhouni's support for the institute and the Bush administration's plans to restrict embryonic stem-cell research and ban therapeutic cloning.

If he does become director of the NIH, Zerhouni will have plenty to keep him occupied besides the politics of regenerative biology. The agency's budget is on course to double from \$13.6 billion in 1998 to \$27 billion in 2003. But since Varmus's departure, many institute directors have left. Some researchers characterize the agency as lacking strategic direction, and being embroiled in infighting over efforts by Tommy Thompson, the health secretary, to centralize control of its work (see *Nature* 415, 250; 2002).

It would be down to Zerhouni — an unknown quantity to most NIH staff and grantees — to restore the agency's coherence and morale. Supporters say that he has the background to do the job. "Medicine is being driven by a technological explosion," says Reed Dunnick of the University of Michigan in Ann Arbor, president-elect of a radiologists' group, the American Roentgen Ray Society. "Someone with his understanding of technology is exactly right for the NIH." ■

Ecologists seek sustainable future

Rex Dalton

An international group of researchers is preparing an agenda on 'sustainable science' for discussion at the World Summit on Sustainable Development, to be held this summer in Johannesburg.

The initiative is led by the International Council for Science (ICSU) in Paris and the Third World Academy of Sciences in Trieste, Italy, and is coordinated by William Clark, an ecologist at Harvard University.

ICSU, the main international federation of scientific societies, was asked by the United Nations to suggest ways of involving science and technology in the summit, which will run from 26 August to 4 September. But there is scepticism about the summit's ability to reach any meaningful conclusions about how science and technology can be applied in poorer countries. Privately, one official involved opines that it will struggle to achieve anything beyond "people in suits sitting around a table discussing options".

But Jane Lubchenco, president-elect of ICSU and an ecologist at Oregon State University, is supporting the effort. "We have to think differently on how to make the transition to sustainability," she says.

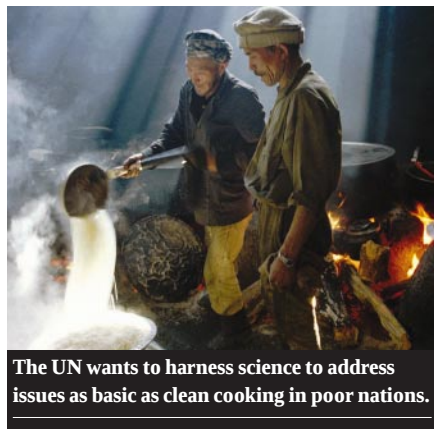
Pamela Matson, an ecologist at Stanford University in California, says that science must become more engaged in addressing the

problems of poor countries. "Development must use water, land and resources in an integrated and sustainable way," she says.

Daniel Kammen, an energy specialist at the University of California, Berkeley, suggests building on established technologies, for example by making clean and efficient kitchen stoves more readily available.

Last week, the initiative's organizers held a workshop in Santiago, Chile, as part of their year-long planning effort. Two more workshops will take place before organizers meet in Mexico City in May to finalize an agenda for the Johannesburg summit. ■

■ sustainabilityscience.org



The UN wants to harness science to address issues as basic as clean cooking in poor nations.

Manchester merger to spawn research giant

David Adam, London

Two of Britain's major research universities are planning to merge in a bid to create an institution that is better able to compete for funds and staff on the international stage.

The University of Manchester and its neighbour, the University of Manchester Institute of Science and Technology (UMIST), announced the proposal on 4 March.

The move would create a university with 28,000 full-time students and a research income of £130 million (US\$185 million) this year — the fifth largest in the United Kingdom behind University College London (UCL), Imperial College in London, and the universities of Oxford and Cambridge (see figure).

UMIST was once formally part of the University of Manchester, but has always had independent governance and financial systems, and the merger plan came as a surprise to researchers at both universities. The plan emerged from a six-month review by a committee of academics at the two institutions into how they could collaborate more closely.

In a joint statement, the respective vice-chancellors of UMIST and the University of Manchester, John Garside and Martin Harris, said: "The proposal to create a new university is a bold and imaginative step. There will now be a full consultation process to discuss the proposal in detail."

The plan could presage more consolidation in UK universities, observers say, although few prospective partners will fit as well as the two Manchester institutions. Their city-centre campuses are barely a mile apart, they share student sports and careers facilities and already have two combined departments, in materials science and civil engineering. If the merger is agreed, it could be implemented by September 2004, university officials say.

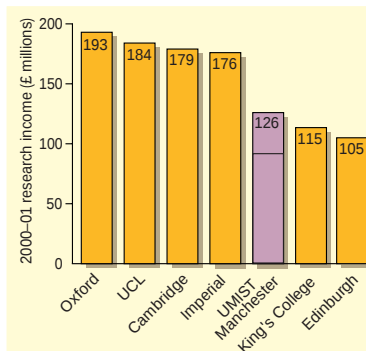
A decade of expansion in student numbers — without a matching growth in funding — has left Britain with 114 universities, most of them relatively small. Several universities have recently swallowed up neighbouring colleges, and some — such as Birmingham and Aston — have discussed merging. "I would be surprised if there were not mergers in the next five to ten years," says Andrew Miller, former vice-chancellor of the University of Stirling. But he predicts that most universities will seek to strengthen their positions through alliances, rather than full mergers.

The universities of Glasgow and Strathclyde, for example, formed a 'strategic alliance' in 1998, through which researchers share equipment, lab space and teaching resources. Looking abroad, Cambridge set up a joint institute with the Massachusetts Institute of Technology in 2000, and Oxford and Princeton announced an alliance last year.

Some university heads, such as Richard Sykes of Imperial College, argue that efforts



Neighbours make nice: Manchester campuses may join to form one of the UK's top research universities.



to compete internationally should be focused on just a handful of UK research universities.

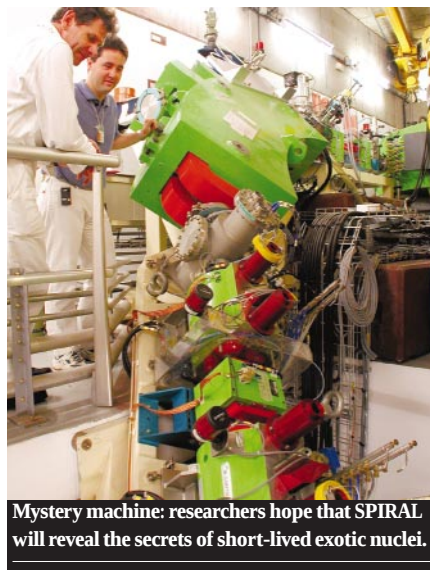
The two Manchester universities stress that their aim is to claim a place at this international high table. But some observers point to devolution in UK domestic politics

as a factor. "It could be that Manchester and UMIST are launching a pre-emptive strike in order to be the dominant partner in a future northwest regional grouping," says John Ashworth, a former director of the London School of Economics. ■

Physicists set sights on exotic prey

Sally Goodman, Paris

A facility for probing the fleeting existence of highly unstable atomic nuclei began operating at the GANIL heavy-ion accelerator in Caen, northwest France, last week. Researchers say that the device, known as SPIRAL, could help to overthrow existing models of nuclear structure.



Mystery machine: researchers hope that SPIRAL will reveal the secrets of short-lived exotic nuclei.

SPIRAL will produce a beam of 'exotic ions' — charged atoms with an unusual balance of protons and neutrons in their nuclei. Most decay into more stable nuclei in fractions of a second. Exotic nuclei are created in extreme environments such as supernovae, and do not occur naturally on Earth.

SPIRAL fires a beam of stable ions at a carbon target held at temperatures of around 2,000 °C. The nuclei of the stable ions fragment on impact, creating a beam of exotic ions. These ions are sorted and then captured by detectors which can, for example, identify them by their decay products.

In most nuclei, protons and neutrons are packed into a ball. But the nucleus of one lithium isotope, for example, seems to consist of a dense centre surrounded by a cloud of neutrons. SPIRAL will aid the search for other such 'halo' nuclei. "We will probably need two or three of this type of facility in Europe to satisfy the demand," says William Gelletly, of GANIL's advisory committee.

Marek Lewitowicz, GANIL's deputy director, says the lab is already seeking to supplement SPIRAL's construction budget of 18 million euros (US\$15.7 million), to boost the range and intensity of its beams. ■

♦ www.ganil.fr

Genomics firm aims to fill Asian gene gap

David Cyranoski, Tokyo

A Taiwanese company last week turned the spotlight on Asian genetic variability when it opened its laboratories in Taipei.

Vita Genomics was formed last year to mine the mass of public and private genetic data in an effort to trace genes among the Asian population that are linked to specific diseases.

Researchers elsewhere in east Asia hope that the company's work will encourage the various governments to increase their efforts to use information on genetic variation to study diseases in the region.

Vita will focus on single nucleotide polymorphisms (SNPs), the single base variations in the genome that account for many genetic differences between individuals. SNPs can be used as markers in studies to identify the genes involved in a population's susceptibility to a given disease and its response to drugs.

By scouring databases of some 5 million SNPs, and comparing them with DNA taken from Taiwanese patient samples, Vita hopes to identify genes that are related to diseases, such as hepatitis, that are common in Asia. According to Vita's president, Ellson Chen, the databases to be used include those main-

tained by the international Human Genome Project and by the US-based company Celera Genomics. Celera, which owns a 5% stake in Vita, helped to set up the company after abandoning its original plan to establish a subsidiary in Taiwan.

"Western drug companies mostly concentrate on diseases common in the West," says Chen. "This has left a gap." As an example, he cites the difficulty of treating hepatitis B and C in Taiwan, where fewer than half of its victims respond to an interferon 'cocktail', the standard drug treatment.

"If we look at SNP patterns on 300 genes which seem to affect whether or not a patient responds to interferon, we could find out why the drug works when it does," says Chen. By July, he hopes to have a first draft of SNPs correlated to some 300 genes related to hepatitis C.

Vita already has a contract with the National Taiwan University Hospital to obtain samples from patients with hepatitis. It also plans to look at patients with breast cancer, liver cancer and asthma. "Even diseases common around the world could have different relevant SNPs in Asian populations," says Chen. Vita intends to cooperate with other institutes in Taiwan, especially the

Academia Sinica, which is greatly expanding its functional genomics capacity.

The question of how genetic diversity between different ethnic groups influences human health is a hotly contested one. But many researchers in east Asia believe that distinct sets of SNPs within a region can influence susceptibility to disease and response to drugs.

Similarities in disease susceptibility across Asian populations have encouraged Vita to mine extensive SNP data that are becoming publicly available in Japan. Yusuke Nakamura, director of the Human Genome Center at the University of Tokyo's Institute of Medical Science, has already found 200,000 SNPs, and has data on how often each variant exists in the population at some 23,000 of these sites (see *Nature* **410**, 1013; 2001).

"These data should help any group studying Asian populations, because the SNPs will behave very similarly among the various Asians," says Pui-Yan Kwok, a geneticist at Washington University School of Medicine in St Louis, who is a scientific adviser to Vita.

◆ snp.ims.u-tokyo.ac.jp

◆ www.vitagenomix.com

Formidable catalogue puts army of ants online

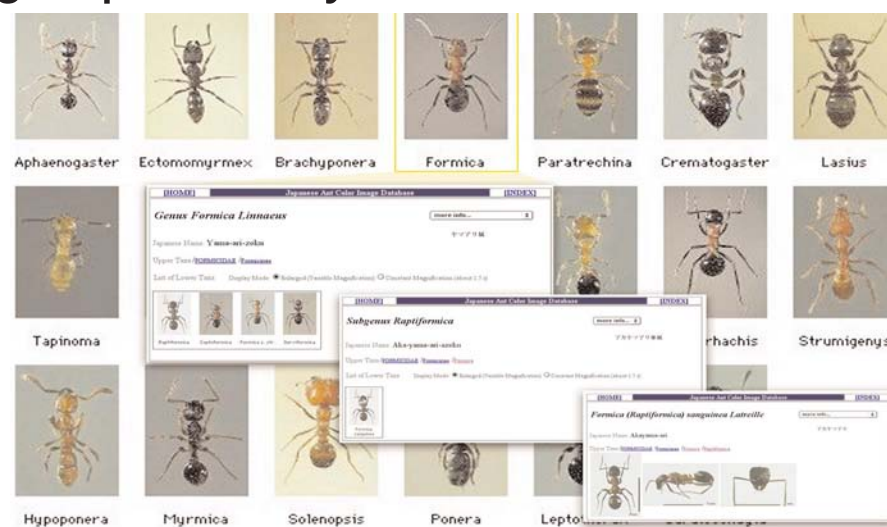
Tom Clarke, London

After four years of cooperation and tenacity worthy of their quarry, ant experts have completed Antbase, a centralized, online information resource cataloguing all 11,000 known species of ant.

Its creators hope that Antbase will one day become the ant equivalent of GenBank, the public database of genetic sequence data, and a boon not just for ant specialists, but for entomologists and ecologists of every kind.

Ant taxonomists Donat Agosti of the American Museum of Natural History in New York and Norman Johnson of Ohio State University in Columbus built Antbase using funds from general research budgets and small institutional grants. Having compiled a full list of ant names, they hope their homespun project will now win long-term funding and a permanent home. "It needs to be institutionalized," says Agosti.

The database's backbone is an updated list of species names. Compiling the list was anything but trivial, Agosti and Johnson say. Descriptions of new species published in obscure journals over the past century had led to heavy duplication. As the project progressed, Antbase shrank to its current size from an initial list of over 20,000 names. "We're trying to pick up the debris from



Ants galore: the foundation of Antbase provides an updated list of species names.

nineteenth-century taxonomy," says Johnson.

Records for each species are being linked to information on their geographical distribution and life history, plus pictures held by the Japanese Ant Color Image Database.

Antbase could emerge as a long-needed resource for ant taxonomists, says Ted Schultz, who works as one at the

Smithsonian Institution in Washington DC. "Before, you had to spend hours photocopying thousands of pages of primary literature," he says. The resource could also aid more general studies of ecosystems, where ants often have a prominent role. "Ants are an incredibly useful ecological indicator," Schultz adds.

◆ www.antbase.org

Institute calls for push to find better anthrax vaccine

Washington The US government needs a broader research effort to produce an anthrax vaccine, according to a report from the Institute of Medicine (IOM).

An expert IOM panel has concluded that the existing vaccine — which consists of a cocktail of proteins from the *Bacillus anthracis* bacterium — is safe, and can protect people against inhalation anthrax. But to achieve protection, it must be administered six times over a period of 18 months, and requires annual booster shots.

The National Institute of Allergy and Infectious Diseases is currently working with the US Army Medical Research and Materiel Command to develop a new vaccine incorporating a recombinant version of the protein that helps anthrax to invade human cells. Clinical trials should begin within three to six months.

The IOM report says that more animal studies are needed to evaluate new anthrax vaccines, and to investigate the effectiveness of combining the current vaccine with antibiotics after the inhalation of anthrax spores.

In 1997, the US military began mass vaccination of its personnel. But more than 400 servicemen and women have refused to take the current vaccine, citing fears that it would cause adverse effects similar to those claimed by Gulf War veterans. The IOM had been asked to study the vaccine to address these worries — but in the light of last year's attacks involving anthrax sent by mail, its report has assumed a new relevance.

♦ www.iom.edu/iom/iomhome.nsf/Pages/Recently+Released+Reports

Pro-life groups denounce stem-cell guidelines

Montreal Canada's main medical research agency has issued guidelines for research on human embryonic stem (ES) cells — prompting complaints from 'pro-life' groups, who argue that the move pre-empts debate on the issue in the federal parliament.

The guidelines, released on 4 March by the Canadian Institutes of Health Research (CIHR), allows federally funded research on pre-existing human ES cell lines, as well as the extraction of new lines from surplus embryos left over from fertility treatments — but only if the parents have given informed consent and no commercial transactions are involved. However, the guidelines ban the harvesting of ES cells from cloned embryos and the creation of embryos for research purposes.

Alan Bernstein, president of the CIHR, rejects complaints that his agency has

subverted the parliamentary process. He says that the guidelines are consistent with draft legislation before parliament, and that they will be amended if necessary.

♦ www.cihr.ca/about_cihr/ethics/stem_cell/stem_cell_intro_e.shtml

It's a virtual blast for nuclear-test computer

Washington The world's most powerful supercomputer has completed the first full three-dimensional simulation of a nuclear blast. The ASCI White computer, located at the Lawrence Livermore National Laboratory in California, modelled an explosion conducted before the United States ended nuclear testing in 1992.

The simulation, which would have taken more than 750 years on a desktop computer, was completed by ASCI White in 39 days. Researchers verified its accuracy by comparing its results to measurements taken during the bomb test.

The successful digital detonation marks an important milestone for the Science-Based Stockpile Stewardship programme, which aims to maintain the US nuclear arsenal in the absence of underground explosive testing.

Combined effort makes optical array a future star

Washington Astronomers have pulled off a technical coup by combining the light captured by six telescopes to produce a single image — the first time this has been done in the visible spectrum. Such an array simulates the performance of a much larger telescope, equal in size to the maximum separation between the different elements.

The Navy Prototype Optical Interferometer (NPOI) on Lowell Observatory's Anderson Mesa site near Flagstaff, Arizona, was used to image the triple star system Eta Virginis with resolution equivalent to that of a 50-metre telescope. According to the project's astronomers, the NPOI array will eventually reach 430 metres in size, which will allow direct imaging of spots on the surfaces of stars.



Array of light: the six telescopes of Lowell's optical interferometer produce one image.

MICHAEL COLLIER/USNO

Catalan deal secures Spanish synchrotron

Madrid The Spanish and Catalan governments last week announced that they will together spend 120 million euros (US\$105 million) on a national synchrotron source to be built in Barcelona. Construction of the 2-giga-electron-volt facility, known as the Synchrotron Light Laboratory, will begin in January 2003, and operation should start by the end of 2008.

The synchrotron is being built for Spanish physicists, biologists and materials scientists to conduct structural studies. But the project's coordinator, Joan Bordas, hopes it could also serve a wider community of researchers across southwest Europe. "It will be the only national facility in this very large area," says Bordas. "All the others are northeast of a geographical line running from Paris to Trieste in Italy." These include the French Soleil facility, which will be built near Paris, to which Spain has agreed to contribute (see *Nature* **409**, 750; 2001).

Scientists leave general public in state of indecision

London Do the public trust scientists? The answer, from a meeting of researchers, politicians, lobby groups and the general public in London on 6 March, was a resounding yes, no or maybe, depending on who you asked.

Issues discussed included a perceived lack of mechanisms to involve the public in debate about key scientific issues, concerns about scientists' commercial conflicts of interest, and the difficulties of communicating science through the media. Nearly 400 members of the public applied for 100 available tickets for the event, which is part of the Royal Society's 'Science in Society' programme, after the society took the unusual step of inviting applications through newspapers and magazines.

The event was held against a backdrop of fierce public debate over issues such as the handling of last year's epidemic of foot-and-mouth disease, and the safety of the combined measles, mumps and rubella (MMR) vaccine.

Fish stock plan nets caviar trade reprieve

Paris States bordering the Caspian Sea have been given the go-ahead to resume trading in caviar after coming up with a coordinated programme for managing stocks of sturgeon. Trading was halted in June last year by CITES, the Convention on International Trade in Endangered Species of Wild Fauna and Flora, after illegal harvesting of caviar — sturgeon eggs — threatened some species with extinction (see *Nature* **412**, 108; 2001).



Fishy business? Illegal harvesting of caviar left some species of sturgeon facing extinction.

The Caspian states account for some 90% of the international trade in caviar, but illegal harvests are around 10 times higher than the legal quotas. The states have now agreed to exchange information on the release of young sturgeon into the sea and to invest more heavily in hatcheries. Natural spawning sites in the rivers feeding into the sea have been decimated by construction projects.

CITES officials are continuing to work with Interpol and local agencies to tackle the bigger problem of illegal trade. But some wildlife organizations are urging CITES to reconsider its decision to resume the legal trade, as the problem has not yet been solved.

Geneticists snap up first photo-booth prizes

New York Dynamite gave us the Nobel Prizes. Now the familiar automatic photo booth has spawned some new awards of similar cash value. The Dan David prizes consist of three annual awards, each of US\$1 million, for achievements concerning the "past, present and future". The awards are administered by the Dan David foundation, based at Tel Aviv University in Israel, and are funded through a \$100-million endowment by David, founder of the company Photo-Me International, which operates some 25,000 photo booths worldwide.

The first winners were announced in New York last week. Three researchers share the inaugural prize in the 'future' category, for their pioneering work in DNA mapping and sequencing: Sydney Brenner of the Salk Institute for Biological Studies in La Jolla, California; John Sulston, former director of the Sanger Institute near Cambridge, UK, and Robert Waterston of Washington University in St Louis.

► www.dandavidprize.com

Picking up the pieces

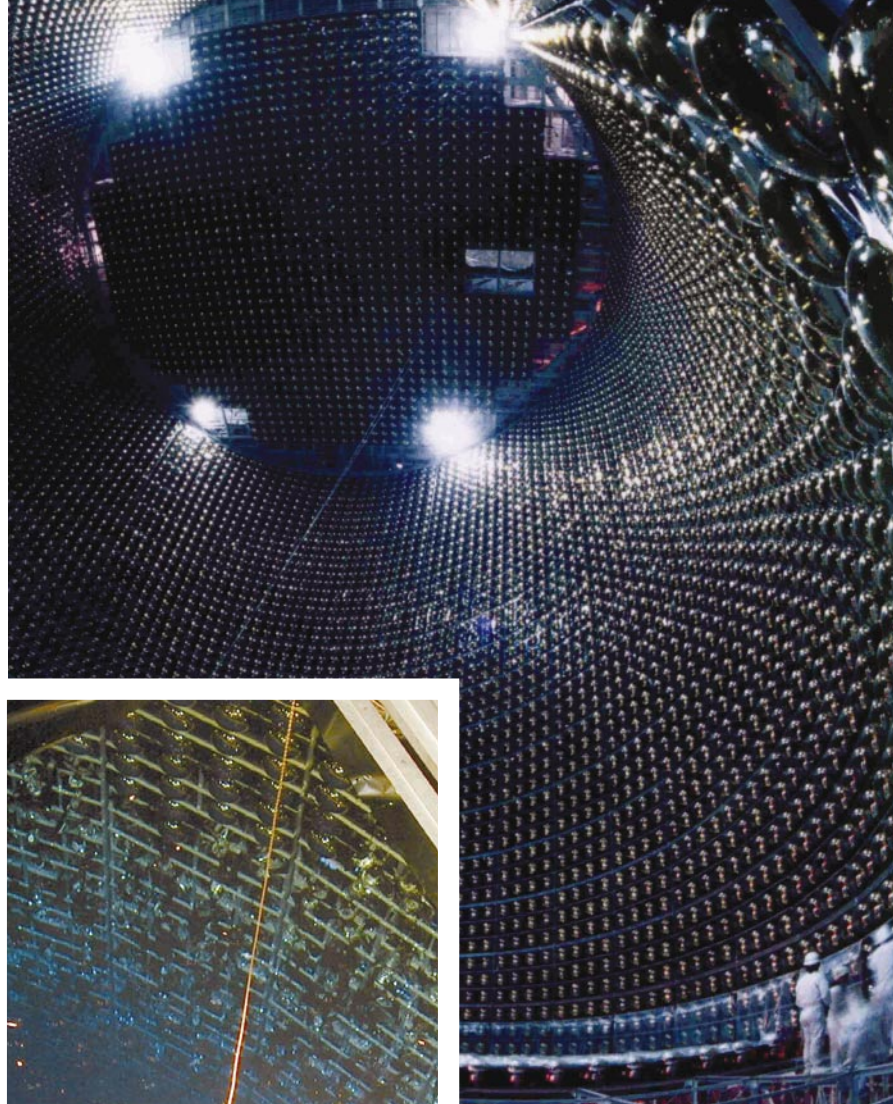
Last November, a shock wave crippled Japan's Super-Kamiokande neutrino detector. David Cyranoski and Geoff Brumfiel find out how physicists plan to resurrect the device.

Noboru Furuta was standing outside a room above the giant Super-Kamiokande detector when he felt a gust of wind. The vinyl flaps covering the doorway behind him blew apart in the breeze. Gurgling and clanging noises filled the air. "All these sounds were mixed together," says Furuta, a technician at Super-K, as the detector is known. "I'd never heard anything like it." The furore picked up pace, like popcorn popping, before coming to an abrupt halt.

Furuta didn't know it, but a shock wave had just ripped through the detector — a huge cylindrical tank filled with 50 million litres of ultra-pure water and lined by some 11,000 light-sensitive sensors. More than half of the sensors had been smashed. In less than 10 seconds, one of the past decade's most significant physics experiments had been reduced to tatters.

Situated in a disused mine below Mount Ikeno, 240 kilometres northwest of Tokyo, Super-K was designed to look for elusive subatomic particles called neutrinos. Collisions between neutrinos and water molecules in the detector's tank produce characteristic flashes of light that are captured by sensitive photomultiplier tubes. The layers of rock above the detector serve to screen out other particles that could produce similar flashes.

Super-K hit the headlines in 1998, when studies there cast doubt on the long-held assumption that neutrinos have zero mass (Y. Fukuda *et al. Phys. Rev. Lett.* **81**, 1562–1567; 1998). The researchers had studied neutrinos produced when cosmic rays collide with the Earth's atmosphere, but in 1999 they sought to confirm their finding by focusing on another source — a beam of neutrinos sent through the Earth's crust from a particle accelerator at KEK, Japan's



Before and after: more than half of Super-K's sensors were destroyed in the accident (inset).

High Energy Accelerator Research Organization in Tsukuba, near Tokyo. This experiment was paused last summer, as researchers drained the tank to replace burnt-out photomultipliers. Disaster struck on 12 November, when the tank was being refilled.

Shattered dreams

The roar of the shock wave, which was picked up by seismometers eight kilometres away, threw those present into a state of panic. In a room close to Furuta, Masayuki Nakahata was monitoring neutrino detections. Nakahata, a physicist at the University of Tokyo, had worked on Super-K's predecessor, Kamiokande, and had been involved in Super-K long before data collection began in 1996. "It felt like it was right under me," he recalls. "Like it was sending spears up through me."

Nakahata's computer registered a dramatic leap in neutrino detections as almost all of the 11,146 detectors lit up before going dark. Thinking the water tank might have burst open, he sent a technician to investigate while he rushed to shut off the power at the four units that house the controls to the detector's high-voltage supply. "I think I

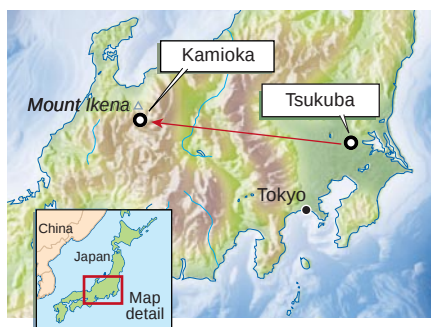
probably went to each one three times," he says. "I didn't know what I was doing."

The technician found no evidence of leaking from the tank, so Nakahata lowered himself on a platform until he was suspended just above the water. The dim light revealed thousands of pieces of black plastic, which normally tethered the black sheeting behind the sensors, floating on the water. The bottom of the tank, usually visible, was obscured by a whitish haze. Twisted wires and the shattered remains of the photomultipliers dangled out of the support frames. Except for those near the surface and above the water, all appeared to be broken. "I knew I had to call and tell people," Nakahata says. "But I couldn't move."

Yoji Totsuka, director of the University of Tokyo's Kamioka Observatory, which runs Super-K, arrived back from a neutrino conference in Canada on the day of the accident.



State of shock: Masayuki Nakahata.



Super-K should soon be able to continue analysis of high-energy neutrinos from KEK's accelerator.

Phoning his wife from the airport, he was told staff at the detector had been trying to contact him. He phoned Nakahata, who gave him the details. "I never dreamt of this kind of destruction," says Totsuka, recalling his shock on learning the news. When researchers assessed the damage over the next few days, they found that 6,779 of the sensors had been destroyed.

Within a week of the accident, a committee of physicists began searching for the cause. The team did not have to look too far. The photomultipliers resemble giant light bulbs, roughly half a metre in diameter. Those at the bottom must each support a huge weight of water. If one was damaged, the committee theorized, the surrounding water could crush it. The resulting shock wave might then shatter neighbouring detectors, creating further shock waves.

Blow out!

Confirmation came when the team lowered nine photomultipliers, arranged in a grid, to the bottom of the tank. An explosive bolt was used to break the central detector, and the resulting shock wave swept through the array, shattering the sensors in a few milliseconds.

The team was also able to trace the course of the accident, as each imploding bulb marked its own destruction with a signal. Two tubes at the bottom of the tank — one of which had been replaced during the refurbishment — were identified as the starting point. The committee concluded that the neck of one of these tubes could have been weakened when technicians walked over them, despite the use of thick foam padding to protect the bulbs. The pressure on the tube would have increased as water filled the tank, eventually causing it to implode.

Even before the investigation was concluded, Super-K researchers made it clear that no individual was at fault. More than 100 physicists worked on the design, but no one had spotted its fatal flaw. "This is one case where you are not going to be able to assign blame," says Francis Halzen, a physicist at the University of Wisconsin in Madison, who works on other neutrino-detector projects.

Totsuka feels differently. He has run Super-K since 1995, and takes very seriously the

Japanese code of honour that requires those in authority to take personal responsibility when things go wrong. Totsuka is due to retire next year, but on 18 February, he offered his resignation as director of the Kamioka Observatory. "We couldn't determine who was at fault, so the person at the top must take responsibility," he says. "Someone must be punished for this huge loss of taxpayers' money."

On 20 March, senior researchers at the University of Tokyo's Institute for Cosmic Ray Research, of which Kamioka is part, will meet to consider Totsuka's resignation. But Motohiko Yoshimura, the institute's director, doubts whether the group will find Totsuka to be blameworthy and assumes that the question of resignation will be dropped.

While colleagues consider his future, Totsuka is focusing on getting Super-K working again. Within days of the accident, he made it clear that the detector should be rebuilt. But full recovery will not come cheap. Around ¥2.4 billion (US\$18 million) is needed to pay for replacement sensors — a big investment at a time when many Japanese science projects are being cut



Yoji Totsuka is determined to restart research.

back or streamlined. The education ministry will not comment on the matter before the University of Tokyo submits its budget request this summer. If the money is forthcoming, the new sensors could be ready in two years.

In the meantime, undamaged sensors and spares are being used to restore Super-K to half strength. Fears that outside workers could be injured if other tubes imploded, coupled with a lack of money, led researchers to do most of the work themselves. Some of Japan's top physicists spent most of February paddling around inside the giant tank removing intact detectors for damage testing. All the sensors will eventually be removed, and each will have a protective plastic case added around the bulb. The sensors will then be fitted back in the tank in a chequer-board pattern, covering alternate spots.

Business as usual

Totsuka hopes to restart the neutrino beam experiment in October, and is confident that the reduced number of detectors can still generate useful data. Other projects may not be so lucky. Super-K's ability to study neutrinos emitted by the Sun, for example, will be severely limited. Neutrinos with energies below a certain minimum level are invisible to the detector, and the initial repairs will leave this lower limit 50% higher than before the accident. High-energy neutrinos from the KEK particle accelerator will not be affected, but some

lower-energy solar neutrinos will now be impossible to spot.

The damage will also be felt in other fields. Another of Super-K's aims is to watch for proton decays. Different theories predict different types of decays for protons, and detecting such a decay would help physicists choose between the competing models. But proton decays are thought to be very rare, and have never been observed.

If a proton inside Super-K's tank did decay, it would produce a characteristic flash of light that could be distinguished from those caused by neutrino collisions. But with the detector operating at half strength, researchers are unsure if they could spot such a rare occurrence. "Our analysis is like looking for needles in haystacks — now it's harder," says Ed Kearns, a proton-decay specialist at Boston University who has worked on Super-K.

Astronomers on the hunt for supernovae could also suffer. The wave of matter thrown out by supernovae slows down light from the exploding star, but does not hinder the burst of neutrinos emitted by the explosion. Super-K was part of an international collaboration, known as the Supernova Early Warning System, designed to alert astronomers to supernovae. But until it is restored to its full capacity, the detector will have to withdraw.

The coming six months will be a critical time for Super-K, as Totsuka awaits a decision on his future, and physicists hope the Japanese government backs full refurbishment. Whatever the outcome, those involved say that they have, in some sense, been fortunate. "We were lucky," says Totsuka. "Lucky that no one got hurt; lucky it didn't happen before we got some interesting data."

David Cyranoski is *Nature's* Asian-Pacific correspondent; Geoff Brumfiel is *Nature's* Washington physical sciences correspondent.

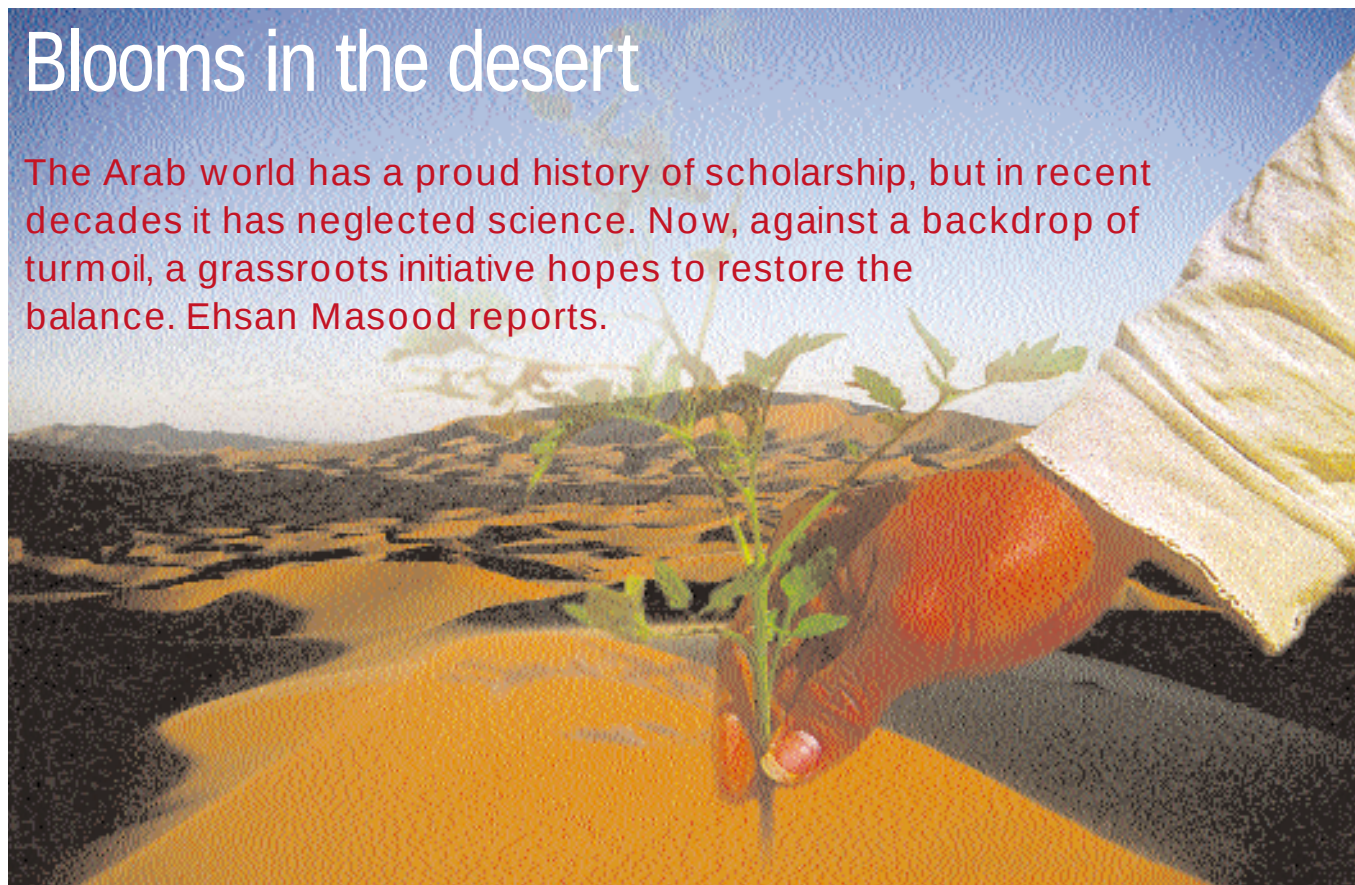
► www-sk.icrr.u-tokyo.ac.jp/doc/sk



Light work: engineers at Super-K replace a sensor during the pre-accident upgrade.

Blooms in the desert

The Arab world has a proud history of scholarship, but in recent decades it has neglected science. Now, against a backdrop of turmoil, a grassroots initiative hopes to restore the balance. Ehsan Masood reports.



M. XERIDATI/L. RUBIN/STILL PICTURES

These are dark days for the Arab world. In many of its nations, standards of living are in decline. The Palestinian-Israeli conflict seems to get bloodier by the day. And further turmoil looms, as the US administration rattles its sabres at Iraq.

But for Arab scientists, at least, there are some chinks of light. Throughout the region, spending on research and development (R&D) is on the rise — admittedly from a very low level. Later this month, scientists will meet in Sharjah, one of the United Arab Emirates, to set the agenda for a foundation that plans to award peer-reviewed grants to the region's researchers. There is also talk of establishing a fund to build scientific capacity across the entire family of Muslim nations.

There is just cause for optimism, says Egyptian-born Ahmed Zewail of the California Institute of Technology, who won the 1999 Nobel Prize in Chemistry. "But countries in the Arab world can and must do much more," he says. "They have both the money and the human resources."

A thousand years ago, Arab science led the world. While Europe languished in the Dark Ages, Arab scholars were pioneering developments in algebra and the study of modern astronomy. But today, the region is — for the most part — a scientific desert. In some states, oil wealth has allowed the construction of fabulous cities, magnificent mosques and sumptuous shopping malls. But little scientific infrastructure has

emerged. Collectively, the Arab nations spend only 0.15% of their gross domestic product (GDP) on R&D, well below the world average of 1.4%.

But in recent years Arab states have realized that the petrodollars won't keep flowing for ever. Already, many of the region's nations are experiencing declines in GDP, and are starting to invest in science in the hope that it will underpin future economic development (see map, opposite).

Fresh focus

Take Oman: between 1992 and 1996, the latest period for which figures are available, its R&D budget increased by more than 83%, whereas its per-capita GDP fell by some 10%. Yemen, meanwhile, managed to increase its spending on R&D by more than 55% over the same period, despite suffering a 65% fall in per-capita GDP. Total R&D

expenditure in the Arab world amounted to US\$782 million in 1996, compared with \$579 million in 1992. In this time, the number of full-time researchers grew by 8%.

Given these trends, leaders of the nascent Arab Science and Technology Foundation (ASTF) hope that they are pushing at a door that is already ajar. Based in Sharjah, the foundation is a coalition of Arab scientists — both those working in Arab nations, and expatriates who have built their careers in the West. It aims to evolve into a pan-Arab funding body that will operate like a Western granting agency — with political independence and thorough peer review. "We want to raise the quality of research in Arab institutions, end the professional isolation of those Arab scientists who have few external links, and eventually infuse research fields with new talent," says Muhammad Garwan, director of the Center for Applied Physical

We're all working scientists. This is not a body of people about to issue resolutions and declarations.

Abdalla Alnajjar



Sciences at the King Fahd University of Petroleum and Minerals in Dhahran, Saudi Arabia.

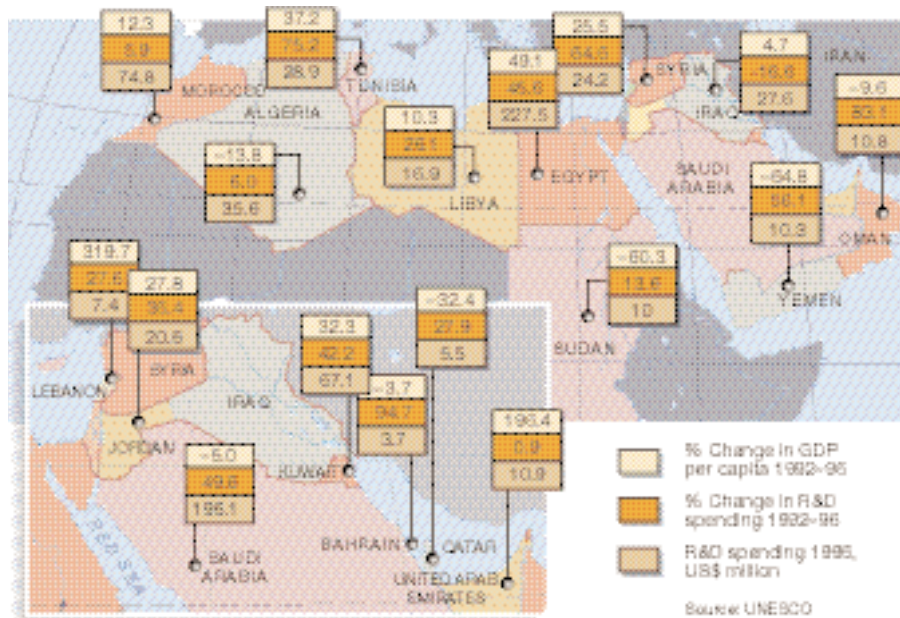
Launched in 2000, the foundation's goal is to nurture blooms in the desert of Arab science. Its progress to date has been underpinned by a US\$6-million donation from the ruler of Sharjah, Sheikh Sultan bin Mohammed Al Qassimi, a graduate in agricultural science. Unusually for a member of a Middle Eastern royal family, Al Qassimi promises to be hands-off, and not interfere in the foundation's day-to-day affairs.

The ASTF's board hopes to raise up to US\$150 million over the next five years. The government of the United Arab Emirates has already pledged support, although it has yet to announce a cash figure. Some 500 Arab scientists are meeting in Sharjah on 24–27 March to draw up priorities for the ASTF. Information technology, materials science, biotechnology, and research into energy and water supplies are expected to feature prominently.

Turning the tide

The ASTF's leaders are anxious not to second-guess the discussion in Sharjah — they want the foundation's priorities to be set by the region's scientists, rather than be imposed by a few influential figures. But they are willing to point to some existing successful Arab scientific initiatives whose quality they hope to match.

The Arab sports of camel racing and falconry have spawned centres with generous funding and strong overseas collaborations.



The Camel Reproduction Center in Dubai, for instance, is among a handful of labs in the world devoted to this field, and has links with the University of Cambridge. The National Avian Research Center in Abu Dhabi is another example of an institution with good overseas ties — it has chosen Carmarthen in Wales as the site for its Falcon Research Institute.

Arab countries also excel in desalination research. Middle Eastern states provide around 50% of the world's desalination capacity. The amount of water produced by this process in Arab countries is expected to double in the next 20 years. But research in

this field is a good example of how international collaborations in science need to navigate the region's troubled politics.

As part of the Middle East peace process, the European Union, Israel, Japan, Oman, South Korea and the United States set up the Middle East Desalination Research Center in Muscat, Oman, in 1996. The centre's mandate is to spearhead collaborations between researchers in Arab countries and those abroad. But deteriorating relations with Israel have meant that researchers from Gulf countries are lukewarm about working with the centre. And US sanctions are preventing

Still open for business

The current 'war on terror' has placed an uncomfortable spotlight on Arabs living and studying in the West. But although fears of a backlash have driven some students away, the long-term prospects for Arab scientists wanting to reach out to their Western counterparts seem brighter.

Arab-Western scientific links have undoubtedly suffered some damage. Mujid Kazimi, a professor of nuclear engineering at the Massachusetts Institute of Technology, says a "sizeable number" of students from the Gulf region left the United States after the terrorist attacks and did not return. Kazimi, who was born in Jerusalem, suspects that more Arab students might want to study closer to home in the future.

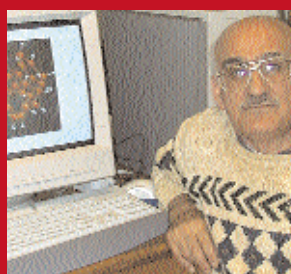
Germany — shocked that Mohammed Atta, leader of the 11 September hijackers, studied at the Technical University of Hamburg — has forced its universities to provide

information about students, including their place of birth and religious background. Student representatives fear the move will fuel anti-Arab discrimination (see *Nature* **414**, 478; 2001).

"It's more difficult for scientists from Muslim countries to find positions in Western institutions," says Atta-ur-Rahman, coordinator-general of Comstech, the ministerial organization that promotes scientific cooperation among Muslim nations. Many postgraduates in the United States, he adds, are returning to their home countries instead of seeking postdoc opportunities abroad.

But scientists of Arab extraction living in the United States report few problems. Kazimi says that he sees no evidence of a reverse brain drain. "Arab and Muslim Americans, as well as those here on permanent visas, are not showing any signs of it," he says.

In the immediate aftermath of



Munir Nayfeh: Arab scientists should have few travel worries.

last year's terrorist attacks, many Arab and Western scientists postponed visits to one another's labs. But things now seem to be returning to normal.

Palestinian-born Munir Nayfeh, professor of physics at the University of Illinois at Urbana-Champaign and president of Arab Scientists and Technologists Abroad, says that his links with Arab labs have been unaffected. He has just returned from

a trip to Saudi Arabia, where scientists were anxious about visiting the West. Nayfeh tried to ease their fears: "I had a visitor in my lab from Damascus on 11 September. I've subsequently had visitors from Saudi Arabia and from Jordan; none have had any difficulty travelling or working."

Nevertheless, experts in immigration law warn Arab scientists visiting the United States to ensure that they have the correct paperwork. Suzanne Brummett, associate attorney with Paparelli and Partners in Irvine, California, says that researchers can expect tougher screening at airports and longer waits for visas. They should also keep their documents with them at all times — for students, this includes evidence of enrolment at a college or university. "Carry proof of immigration status," says Brummett. "If you don't, you could face 30 days in prison or a \$200 fine."

participation by researchers from Libya.

Building and maintaining links with Western labs will be crucial to the ASTF. A significant part of the foundation's budget will be earmarked for research collaborations between scientists in Arab nations and those outside the region.

To establish such links, the ASTF is looking particularly to Arab researchers who have made their careers in the West. Shadia Rifai Habbal, a solar-terrestrial physicist at the University of Wales at Aberystwyth, is one such scientist. The ASTF presents "tremendous opportunities for both sides", she says. Habbal, who was born and educated in Syria, studied the 1999 solar eclipse with the help of researchers in that country.

There's an unmistakable sense of optimism among the ASTF's members. But talk to some of the older hands, and the magnitude of the task the ASTF is taking on becomes clear. Arab science has experienced several false dawns, as expert committees drew up strategies that have subsequently gathered dust on the shelves of the officials charged with their implementation.

Antoine Zahlan, formerly a professor of physics at the American University of Beirut and author of several books on Arab science policy, knows these frustrations first hand. Zahlan, who now lives in London, helped to draft a science strategy for the Arab League nearly two decades ago, which called on the league's members to invest at least 1% of their GDP in research. Few governments implemented any of its recommendations.

More than just a building

"The problem isn't necessarily a lack of money, or buildings, or trained PhDs," Zahlan says. Rather, there has been a failure to understand science's role in society — and to realize that there's more to building scientific capacity than opening a gleaming new lab. Arab scientists complain that budgets are consumed by salaries and administration, leaving little for research. Experts also warn that official statistics on increases in science spending must be interpreted with caution, as these may include investments in higher education. "Even when you want to do research, the funds simply aren't there," says one senior Arab scientist, who asked not to be named.

Nader Fergany, director of the Almishkat Centre for Research in Cairo, an independent social-sciences research organization, blames the autocratic nature of many Arab societies for stifling scientific inquiry. "The prevailing social system is paternalistic, based on submission to, and fear of, authority," Fergany says. "Most Arabs lack the rudiments of functional literacy. An even greater majority has lost the ability to think critically and express themselves freely, and almost all are barred from effective participation in public affairs."

In such an environment, researchers tend

Science veiled in secrecy

Researching an article on science in the Arab world can be frustrating. Scientists in many Arab countries are reluctant to speak to foreign reporters on the phone. Correspondence by e-mail — which is easily monitored — is discouraged. Questions usually need to be faxed in advance. And don't bother asking about the repercussions of the events of 11 September 2001 — such enquiries will most probably be politely rebuffed. Ordinary jobbing scientists are afraid of saying anything that might upset the ruling authorities.

This culture of self-censorship has also hampered efforts to quantify the research activity of Arab countries. Even where information is willingly supplied, it is difficult to assess its accuracy. "In a few countries, the machinery to compile and analyse data simply does not exist," says one UN official. "They will just guess at the numbers and jot down what they think is right."



Omar Bizri sees a long road ahead.

Were it not for the persistence of Adnan Shihab-Eldin, former head of UNESCO's Cairo office, and Omar Bizri, head of the technology section at the UN Economic and Social Commission for Western Asia (ESCWA), based in Beirut, this article would be bereft of data. These two individuals were behind a 1998 study on science indicators in the Arab states, which relied heavily on dispatching emissaries to ministries and research labs with letters of recommendation to personal contacts. "In the Middle East, personal contact has always been more effective than formal contact in all transactions," says Shihab-Eldin, who is now director of research at the Organization of the Petroleum

Exporting Countries in Vienna. "It is a cultural feature."

To make future reports more reliable, ESCWA and UNESCO, with the help of the Organisation for Economic Co-operation and Development, aim to set up data-collection teams in each country, with properly trained personnel. Definitions of, for example, what constitutes spending on research and development also need to be agreed. At present, says Bizri, different countries have different interpretations. "A lot more work needs to be done."

to become inward-looking and wary of outsiders (see 'Science veiled in secrecy', above). Similarly, some funding agencies in Arab countries are reluctant to fund collaborations with foreign researchers, fearing that their governments will accuse them of leaking sensitive data to the West.

The ASTF's founders acknowledge the difficulties, but argue that they have already broken the mould by gathering together active researchers who understand the need for open scientific interchange. The foundation's origins lie in the frustrations of Arab scientists who want to do better research, says ASTF president Abdalla Alnajjar, a physicist and director of research at the University of Sharjah. "We want to be practical," he says. "We're all working scientists. This is not a body of people about to issue resolutions and declarations."

If the ASTF can engender a new culture of scientific openness in the Arab world, Western agencies seem likely to respond in kind. So far, most Western countries have few bilateral science agreements with Arab countries. But there is a desire to build stronger links.

Positive action

The United States has just one such agreement — with Egypt — in which \$2 million a year is allocated through the US Agency for International Development to Egyptian scientists for joint activities with US-based researchers. William Gaines, director of the Office of Science and Technology Cooperation in the State Department, hopes to increase the funds available for this collabora-

tion to \$3 million, and to sign agreements with other Arab states. "Preliminary indications are positive," he says. Similarly, officials with the European Commission in Brussels are keen to use scientific projects to help forge closer links with Arab nations.

At the same time, there are also rumblings of the creation of a new research fund to foster scientific links among Muslim countries. Announced at the annual meeting of science ministers from the Organization of Islamic Countries (OIC) in Islamabad, Pakistan, last month, the precise scope of the fund remains unclear — but potentially, it could dwarf the ASTF. The fund is the brainchild of Atta-ur-Rahman, coordinator-general of Comstech, which supports scientific cooperation among the OIC's member states. Rahman has a reputation for getting things done: last year, as Pakistan's science minister, he oversaw a 50-fold increase in his ministry's budget for R&D projects.

Although it is too early to predict the impact of the ASTF and Rahman's proposed fund, it does seem that Arab science is at last beginning to emerge from its years in the desert. None too soon, say researchers, given that the region's future may depend on it developing a knowledge-based economy. "Countries that want to progress need to invest in their human resources," says Rahman. "The real wealth of a country is its people."

Ehsan Masood is director of communications at the sustainable development network LEAD International, based in London, and was formerly a reporter for *Nature*.

♦ www.astf.net

In risk assessment, one has to admit ignorance

Explaining there are things we can't know could improve public confidence in science.

Sir—The risks and benefits of novel technologies such as genetically modified foods continue to be fiercely debated. Risk assessment is at the crux of these conflicts, as shown, for example, by the report of the United Kingdom's Agriculture and Environment Biotechnology Commission last September, *Crops on Trial* (www.aebc.gov.uk/aebc/reports.html). Yet it is often overlooked that scientific risk assessment is fundamentally limited by ignorance.

Naturally, scientific knowledge refers to known processes and their influence upon known state-variables. Within this domain of reproducibility and control, uncertainty can be explicitly stated and reduced by reproducible experiments under controlled conditions. However, the domain of ignorance, characterized by the interaction between unknown processes and/or unknown state-variables, tends to be implicitly neglected in risk assessment. As the well-known examples of dichlorodiphenyltrichloroethane (DDT) and chlorofluorocarbons (CFCs) suggest, our ability to assess novel risks is primarily limited by the fundamental difficulty of taking these interactions into account.

In the case of DDT, the increase in concentrations of this insecticide resulted in (among other things) fragile egg-shells, threatening the survival of rare species of birds. Here we are dealing with interactions between a known process (increase in DDT concentration) and an unknown, thus neglected, state-variable (egg-shell thickness); between this neglected state-variable and a neglected process (population dynamics); and between this neglected process and the neglected state-variable of bird population. All of these interactions fell within the domain of ignorance for the contemporary environmental risk assessments of DDT. As long as egg-shell thickness was not understood to be a relevant state-variable, there was no reason to monitor it. Therefore, it was exceedingly difficult at the time to identify the unknown effect of bioaccumulation.

In the case of CFCs, risk assessment was initially limited to human toxicity. The highly stable nature of CFCs was considered desirable because it was thought to indicate the very low reactivity of these novel compounds and thus their suitability as non-flammable refrigerants. Vertical transport of CFCs into the stratosphere was not then considered a relevant process; their concentration in the stratosphere was not monitored (neglected state-variable). Nobody suspected a connection between stratospheric CFC

concentrations and stratospheric ozone concentration (neglected photochemical processes in the stratosphere). Once again, these state-variables and interactions were neglected because they belonged to the contemporary domain of ignorance.

The impossibility of taking unknown processes and variables into account may be a more fundamental obstacle to credible risk assessment than our inability to describe the known interactions accurately. Yet the current discussion on uncertainty tends only to deal with the latter.

The possible consequences of ignorance are a major concern among the public regarding new technologies: time and again, people ask who will be in charge of responses to inevitable future surprises, and whether they can be trusted.

The policy response to these concerns has typically been to recommend further research on known uncertainties, with the intention of creating greater certainty and hence reassurance that risks are controlled. This response fails to reassure the public since it mistakenly assumes their concerns to be inspired by a demand for zero risk.

To overcome mutual misunderstanding by scientists, policymakers and the public, it is important for all to acknowledge that unanticipated effects of novel technologies are not just possible but probable — and that potential harmful consequences cannot reliably be established by further research since they fall into the domain of ignorance.

Risk assessment and policy need to

emphasize uncovering the limits to knowledge, rather than proving existing knowledge to be correct. Multiple interacting perspectives should be encouraged, as each can be useful in pointing to limitations of the others. Lay knowledge, in particular, can be a valuable addition to expert knowledge, because it is based on different experiences.

Both the UK Office of Science and Technology guidelines on scientific advice for government (www.dti.gov.uk/ost/aboutost/scientific_advice/index.htm) and the European Union communication on the precautionary principle (europa.eu.int/comm/dgs/health_consumer/library/pub/pub07_en.pdf) laudably advocate greater inclusiveness, transparency about uncertainties, and accountability in scientific inputs to risk policy. But they also need to recognize and address the crucial distinction between uncertainty and ignorance. Otherwise, they may inadvertently contribute to the continuing confusion of the pretence of control with the reality of unanticipated consequences. Failure to address this predicament may unintentionally encourage further erosion of public confidence in science.

Holger Hoffmann-Riem*, **Brian Wynne†**

**Institute of Terrestrial Ecology, Swiss Federal Institute of Technology, Grabenstrasse 3, 8952 Schlieren, Switzerland*

†Institute for Environmental Philosophy and Public Policy, Furness College, Lancaster University, Lancaster LA1 4YG, UK

Favouritism in physics?

Sir—In his Correspondence (*Nature* **415**, 472; 2002), a response to your Opinion article (*Nature* **414**, 829; 2001), L. Maiani states that, in comparing CERN with DESY, “you fail to take into account the significant cost-of-living differentials between Geneva and most other European locations”. Most CERN employees do not suffer from the high cost of living in Switzerland because they live and do most of their business a few kilometres away in France. They get the best of both worlds: high salary with excellent benefits, and a relatively low cost of living. In this sense, if no other, they are ‘more equal than others’.

For decades, particle physics in Italy has been very well funded. It has been able to attract excellent students and researchers with the prospect of an exciting career, opportunities for travel, and salaries even while studying for their *laurea* projects (equivalent to a master's degree), in stark

contrast with those in other disciplines. This dominance of particle physics is probably due to the legacy of Enrico Fermi, who started the school that became Italy's national institute of nuclear physics, of which Maiani was president before becoming director general of CERN.

Thanks to this institute, which receives some 250 million euro (US\$216 million) a year from the Italian government, and to the roughly 100 million euro a year that Italy contributes to CERN, Italian particle physics has been receiving more research money — by orders of magnitude — than any other Italian basic scientific discipline. (The Italian research budget is a meagre 1% of its GDP.) This perhaps helps to explain why Italian research has higher standards in particle physics than in other scientific disciplines.

Federico Rosei

Institute of Physics and Astronomy, University of Århus, 8000 C Århus, Denmark

Lots of peanut shells but no elephant

A controversial account of the discovery of HIV.

Science Fictions: A Scientific Mystery, a Massive Cover-up, and the Dark Legacy of Robert Gallo
by John Crewdson
Little, Brown and Company: 2002. 672 pp.
\$27.95

Jaap Goudsmit

"The French isolated the first AIDS virus ever (HIV-1 Lai), and the Americans isolated the first virus that was truly representative of the Western AIDS epidemic (HIV-1 MN). The original Bru [AIDS virus] was even more representative, due to its slow-growing properties, but nobody — not even Montagnier — realized this virus existed until years later." This is what I wrote some five years ago in *Viral Sex: the Nature of AIDS* (Oxford University Press, 1997) about the controversy over whether human immunodeficiency virus (HIV) was discovered by the French scientist Luc Montagnier or the American Robert Gallo. John Crewdson's book didn't change my mind.

Science Fictions is not exactly bedtime reading, and doesn't add to the previously published accounts of events, despite the fact that new interview material is included. Contrary to the book's subtitle, Crewdson doesn't unveil a scientific mystery, he doesn't unearth a massive cover-up, and even as a biographer of Gallo he fails to provide a clear picture of either his subject's character or his scientific accomplishments.

Crewdson's claim to fame in the AIDS community is his 50,000-word account of the discovery of HIV, and the controversy surrounding it, which was published in the *Chicago Tribune* on 19 November 1989. *Science Fictions* elaborates on the same theme. It ends with the statement: "But every road down which Gallo had searched for an AIDS cure had been a dead end", even though in the book important discoveries are attributed to Gallo — the discovery of the virus HTLV-1 and T-cell growth factor. Crewdson seems determined to convince us that Gallo is a villain, but he didn't convince me.

Three points of interest come to mind surrounding the discovery of HIV and the subsequent events. First, who discovered HIV: Gallo, Montagnier or somebody else? Second, how did the discovery come about and who played what role? And third, did the political, social, economic and psychological warfare surrounding the issues of credit and credibility in any way slow down progress in AIDS research?

The answer to the first question is straightforward: HIV was discovered by Françoise Barré-Sinoussi in Montagnier's laboratory at



Burying the hatchet: Montagnier (right) and Gallo (left) at an awards ceremony in Spain in 2000.

the Pasteur Institute in Paris, with the support of a team of French clinicians and researchers, including Jean-Claude Chermann, Willy Rozenbaum, David Klatzmann and Montagnier. They published their findings in May 1983, about a year ahead of anyone else (*Science* **220**, 868; 1983).

There seemed to be little to make a fuss about. Evidence of a new virus, Bru, present in the lymphoid tissue of a homosexual man, was first obtained in January 1983. Bru multiplied only in primary cell cultures, and with hindsight we know it was a type of virus (non-syncytium-forming, or NSI) that is not transmissible to permanent cell lines.

But the cultures of HIV Bru were contaminated at the Institut Pasteur at some point in 1983 with HIV-1 Lai, a virus from an AIDS patient that multiplied very quickly in cell culture. This syncytium-forming (SI) virus was easily transmissible to permanent cell lines, as Gallo and Mikulas Popovic discovered. All cultures subsequently sent out from the Pasteur, including those sent to Gallo and later to Robin Weiss in Britain, were a mixture of the slow-growing Bru virus and the fast-growing Lai virus. It was no surprise then that both Gallo and Weiss succeeded in

selecting the Lai virus while establishing permanent cell lines from the primary culture they received; this is the prime characteristic of SI viruses. Weiss subsequently supplied his cell line to Montagnier's group, and this marks the start of the controversy: both Gallo and Montagnier had cell lines in which the Lai virus predominated.

We have to credit Birgitta Asjo and colleagues (*Lancet* **2** (8508), 660–662; 1986) with the discovery that slow-growing, low-titre viruses are prevalent in early-stage HIV infection, and that rapidly growing, high-titre viruses are common in late-stage infection. Subsequently, in the early 1990s, several researchers identified the slow/low viruses as NSI viruses and the rapid/high ones as SI viruses.

In 1995–96 the scientific basis of these findings was unveiled, and Gallo did play a pivotal part in this. The story started when Jay Levy claimed that cellular antiviral factors circulating in the body were able to block HIV infection. In December 1995, Gallo and co-workers identified the cellular molecules RANTES, MIP-1 α and MIP-1 β as major HIV-blocking factors (*Science* **270**, 1811–1815; 1995). Gallo credited the first

and last authorship to the people who did the work in his lab — Paolo Lusso and Fiorenza Cocchi — with full awareness of the importance of the paper, as Lusso recently told me.

The only cell receptor on CD4⁺ T cells that binds to these three ligands is CCR5, so it took only a few months to prove formally that CCR5 is the co-receptor for NSI HIV strains, particularly in combination with the finding that CXCR4 is the co-receptor for SI HIV strains. These final parts of the puzzle explain in full the biological background of the emergence, during *in vitro* propagation, of the HIV Lai strain from a mix of SI and NSI viruses. This story is absent from *Science Fictions*, and other parts of the book also seem to have been written blindfolded.

The best illustration that the field did not suffer from the Gallo–Montagnier controversy is the story of the development of blood tests for the HIV virus, based on the detection of antibodies against HIV proteins. To cut a long story short, Crewdson claims that the Pasteur's HIV antibody test was 'better' than the tests originating from Gallo's lab, but that Gallo's 'tests' were approved more quickly and more easily, with the result that supplies of donated blood were less safe. I was personally involved in evaluating blood tests in these early days, and find Crewdson's account of events to be biased.

In blood from AIDS patients, the tests from Abbott Laboratories and Litton Biogenics — based on Gallo's production cell line — were as sensitive as the Pasteur test, and the test from Wellcome was significantly more sensitive than the Pasteur test (*Lancet* 2 (8505), 483–486; 1986). All the tests were significantly less sensitive than the reference standard when we tested blood from symptom-free HIV-infected individuals. We now know that this is because, in the early stages of infection, antibodies against one HIV antigen (p24) predominate, and for technological reasons many of the tests were deficient in picking up p24 antibodies.

Crewdson is right to point out that the Abbott Laboratories test showed high levels of false positives (resulting in some wastage of uncontaminated blood), but other tests based on Gallo's cell line showed as few false positives as the Pasteur test, a point Crewdson has ignored. Finally, in 1986 my colleagues and I showed (*Lancet* 2 (8500), 177–180; 1986), in collaboration with Abbott Laboratories, that p24 is found in the blood before antibodies against this protein appear — a finding that eventually improved the HIV tests substantially — and that the drop in levels of p24 antibodies in the blood of AIDS patients is directly related to the re-emergence of high levels of p24 in the blood (antigenaemia). As soon as we had solid proof that p24 antigenaemia heralds immunodeficiency and AIDS, we started treating those p24-positive individuals at highest risk for AIDS with AZT. None of this

sequence of events indicates that the fight against AIDS was slowed down because of the Gallo–Montagnier controversy.

Too often, Crewdson's book ignores what is obvious: that, in the end, the science counts rather more than the scientist. Particularly when every day 15,000 people are newly infected by HIV. ■

Jaap Goudsmit is at *Crucell and the Academic Medical Center, University of Amsterdam, 1105 AZ Amsterdam, The Netherlands.*

Bare bones of a life

Dart: Scientist and Man of Grit
by Frances Wheelhouse &
Kathaleen S. Smithford
Transpareon, Australia: 2001. 361 pp. £23
C. K. Brain

There have been many important milestones along the way to our ever-deepening understanding of human origins, but one was of special significance. It was Raymond Dart's interpretation, published in *Nature* on 7 February 1925, of the fossilized skull of a child, half-ape, half-human, that came into his hands without warning from a limestone cave at Taung in South Africa. It was the first of its kind ever to have existed. How this 32-year-old anatomist had acquired the necessary neurological expertise to draw his radical conclusions, and the confidence to challenge accepted concepts of his time, is one of the themes of this biography. It is essentially a tribute, devoid of criticism and motivated, it seems, by admiration and respect. The result could easily have proved tedious, had not Dart so richly deserved the kindly treatment given to him by these two authors; instead, it is lively and enthralling.

Coming from a rural Australian background, Dart graduated in medicine at the

University of Sydney. At the end of the First World War, he joined Grafton Elliot Smith in his anatomy department at University College, London. Here Dart's preoccupation with neurology deepened, and received a further boost when he spent a year at medical institutions in the United States on a Rockefeller Foundation fellowship. In 1923 he moved to Johannesburg as professor of anatomy in the Medical School of the University of the Witwatersrand.

Hoping to establish an anatomical museum in his department, Dart encouraged his students to bring him interesting specimens, and in this way acquired a fossilized baboon skull from Taung. He immediately arranged for further fossils to be sent from the lime-works there, and to his delight these included the child skull that he named *Australopithecus africanus*. Apart from the complete face and lower jaw, this specimen had a remarkable endocranial cast, showing the pattern of structures that had existed on the surface of the brain. Dart concluded that this pattern, together with other anatomical features of the skull, indicated that *Australopithecus* had represented an upright-walking link between the apes and humans in Africa.

Although Darwin had speculated 50 years earlier that Africa had been the continent of human origin, discoveries elsewhere had swung scientific attention away from Africa to Europe and the Far East. Furthermore, it was surmised that our early ancestors had acquired large brains before they walked upright. So one can imagine how meagre was the support given to Dart's claims by the scientific establishment. At best, the skull was dismissed as that of an anthropoid ape, of little interest until adult specimens became available.

But in South Africa, unwavering acceptance of Dart's interpretation came from Robert Broom, who described the first of many adult *Australopithecus* fossils from the Sterkfontein cave in 1936. These and subsequent discoveries vindicated Dart's heretical claims and, for many years, Dart was happy to leave the action to Broom. However, one of Dart's best-known students, and his successor in the anatomy department, Phillip Tobias (author of the foreword to this biography), coaxed Dart back into the early-hominid field by showing him the fossil potential of the Makapansgat Limeworks cave. Dart started field investigations there in 1946 and turned up several *Australopithecus* specimens among a vast collection of other mammal bones. Speculating on how these bones came to be in the cave, Dart concluded that the ape-men had been mighty hunters, bringing back those bones that would be useful as tools and weapons, and practising an 'osteodontokeratic culture'.

In a long series of publications, Dart developed his concept of "the predatory transition from ape to man", in which our ancestors "slaked their ravenous thirst on the



Granddaddy: Dart, aged 91, with the Taung skull.

hot blood of victims and devoured livid, writhing flesh". When I asked Dart why he used such powerful prose in his serious scientific writing, he replied simply: "That will get 'em talking." And it certainly did. In my case, it provoked me to spend many years developing the new discipline of African cave taphonomy, by which the origin of such fossil assemblages may be interpreted with confidence. The result was that many of Dart's dramatic concepts gave way to more realistic ones, and the "mighty hunters" were seen to have been "the hunted". As these alternative ideas emerged, I discussed them all with Dart. To my great relief, he was delighted, saying: "This is wonderful — at last we are getting closer to the truth!" He immediately nominated me for an award.

Dart was clearly more interested in the subject than in his own position relative to it. His remarkable generosity of spirit was refreshing in the emotion-driven field of palaeoanthropology, and remains an icon for us all. As is so well portrayed in this biography, his main concern was for his fellow humans, both living and long dead. ■
C. K. Brain is at the Transvaal Museum,
PO Box 413, Pretoria 0001, South Africa.

The shape of things to come

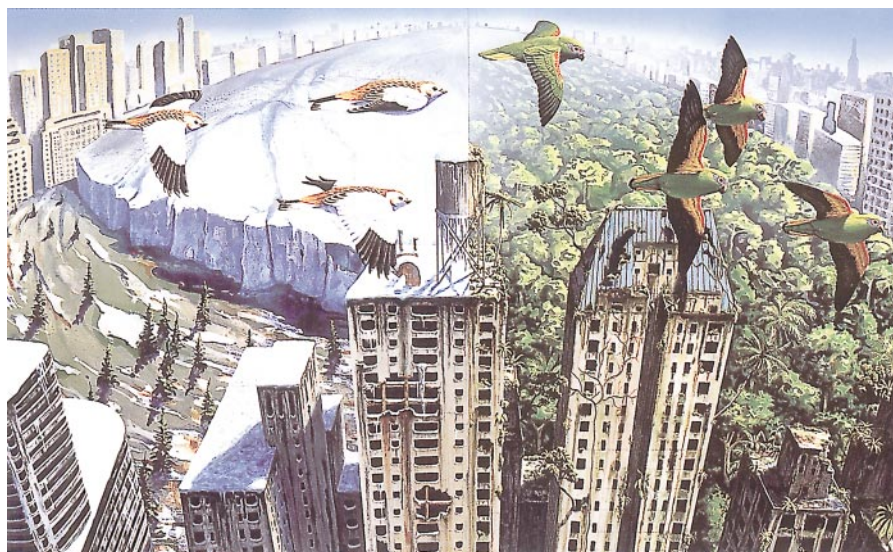
Future Evolution: An Illuminated History of Life to Come

by Peter Ward & Alexis Rockman (illustrator)
Times/Henry Holt: 2002. 208 pp. \$35

Dougal Dixon

Future evolution — what an exciting prospect! Animals and plants, fundamentally changed, populating the Earth until the end of its history. Biological diversity stretching out forever. Such is the promise of this work — and with it comes the prospect of mind-expanding spectacle, and the controversy that stalks any work of futurology.

Peter Ward lays out his arguments in his introduction, summarizing them in eight points. (1) Mass extinctions in the past instigated biological innovation — I can't argue with that. (2) Past extinctions have had several causes — that's OK too. (3) The Earth has been undergoing a mass-extinction event since the end of the Ice Age — most would agree with that. (4) This mass extinction is different — well, maybe. (5) All mass extinctions have been followed by a recovery — that's fine. (6) New species will evolve — indubitably. (7) Our species, *Homo sapiens*, is extinction-proof — now that's contentious. (8) There will never be another dominant fauna — well, that depends on item (7), and is just as dubious.



Which future? Alexis Rockman's double vision reflects the unpredictability of climate.

Let's take a look at his arguments. The first third of the book details past mass extinctions. He provides dates, statistics and specific examples, but most of this is already known. It is well summarized, but is not really what we want from a book like this.

The second section is all about the current mass extinction. The 1.6 million named species today may represent a mere 3% of the total. Did you know that this is more species than existed at any time in the fossil past? It is not clear what he is using as proof, but his logic is sound: the break-up of the 'supercontinent' Pangaea led to many areas of isolation where new species could develop. Modern travel has effectively reunited Pangaea, from the biological viewpoint, and subsequent competition has eliminated many species. This, of course, along with environmental degradation, pollution and all sorts of other ills, is our fault. Again, this is familiar stuff and not really contentious. But we're interested in the future, not the past. That is why we picked up the book in the first place.

This is where Ward becomes a little more controversial. The current mass extinction, he states, is almost over. It is now 10,000 years since the end of the Ice Age, and the mass extinction has stabilized. It was never as bad as the tabloids made out anyway — however many of the 1.6 million species were wiped out, there are still a heck of a lot left. Humans will never become extinct, says Ward, and whatever evolutionary trends take place from now on will have our descendants at their core. Such anthropocentric optimism is never explained, but given that, what kind of future does he foresee?

He launches into a delightful description of "zeppelinoids" — whale-sized animals that float on internal gasbags and trail tentacles for prey. This is more like it. This is the leap of imagination that we were expecting. Alas! It is a negative example. Something that will not happen. His message in this is that

there will be no new body plans. Instead, we have a world in which convergent evolution rules, a world of descendants of modern survivors — the animals that *Homo sapiens* encourages to survive by selective breeding and domestication, and the pests that will survive despite human effort. So we have a future based on the transgenic offspring of pigs, and the garbage-feeding descendants of flies, rats, raccoons and fleas.

What of mankind itself? Well-worn science-fiction themes are explored, and to some extent accepted. There is organic engineering, with industrial materials being grown rather than mined and processed. There is the fact that, through medical science, natural selection has stopped for us. The manipulation of the human genome is acknowledged, although the word 'eugenics', with all its negative overtones, is sidestepped and the moral implications ignored.

Future Evolution is well written. Each chapter begins with a personal anecdote that has some bearing and then develops into the meat of the argument. Now and again Ward climbs into H. G. Wells' time machine and jumps to a point in the future for a look around. But these trips are achingly few, and only last a couple of pages each. The final chapter, the one in which we really hope to see some Hollywood-style computer-generated imagery of the mind, begins with Ward's experiences studying the nautilus off New Caledonia (his day job) and then continues in that vein, rounding off with one of his short, tantalizing visions of the future. It leaves us with an appetite for more — much more.

We can sympathize with Ward's dilemma. He must keep his imagination on a tight rein so that the science behind his ideas is not compromised. The scientific groundwork is well laid, but the imaginative interpretation is thin. Pity! ■

Dougal Dixon is a science writer based in Wareham, Dorset, UK.

The barest essentials

Massimo Piattelli-Palmarini

Two styles of explaining the science of mind and behaviour have been competing for as long as anyone cares to remember: empiricist, centring on habit formation, statistical learning, imitation and association; and rationalist, focusing on the projection of internally represented rules. Despite relentless effort, the former has delivered rather meagre results, whereas the latter, with its pivotal concept of an internally represented grammar, has produced the solid 'conceptual cognitive revolution'.

For a rationalist cognitive scientist, a grammar is a finite mental object, systematically assigning abstract structures to all the well-formed expressions of a language — that is, to each member of a set that, for natural languages (such as Chinese or Italian), is infinite and discrete. Infinite, because every speaker of a language can produce and understand an unlimited number of new grammatical sentences. Discrete, because continuous modification of a sentence to change it into another is impossible. No sentence could be halfway between "It's a good car, but they don't sell it" and "It's a good car, but they don't tell it."

A grammar capable of generating complex structures for all well-formed sentences of a natural language must have recursive rules, because phrasal constituents can contain other phrasal constituents of the same or higher kinds ("The young doctor's three beautiful sisters" is

a noun phrase containing another noun phrase; "The spy who came in from the cold" is a noun phrase containing a sentence). Moreover, structural rules of sentence formation can be applied recursively to embed relative clauses embedding other relative clauses, without limit (as in "This is the cat that killed the rat that ate the malt that lay in the house that Jack built"). Because such grammars are finite, whereas the languages they generate are infinite and contingently shaped by use, it is advantageous, and methodologically cogent, to consider the concept of grammar as primary, and that of language as derived.

Since the mid-1950s, powerful formal criteria, derived from analysis of the artificial languages of mathematics and computer programming, have been applied to the study of natural languages to determine principles by which a given class of grammars can generate a given target language. A universal ('Chomsky') hierarchy of grammars (automata) was established: the most powerful class contains as a subclass the immediately less powerful one, and so on. In tune with the dominant empiricist-inductivist tradition of the 1950s, the first grammars to be explored at the lowest level in the hierarchy were probabilistic and finite-state. From a very large corpus of ascertained utterances of the language, one can compute the conditional probability that a word (or string of words) will follow another.

As well as being utterly unrealistic as cognitive models for a human brain — mind because of the burden on archival memory and the relentless updating of probabilities, such a grammar would be inadequate because long-range syntactic dependencies would be assigned a vanishingly small probability, contrary to fact. A simple example is the agreement between subject and verb in the sentence "The children are laughing", and how it can be interrupted by another agreement relation, as in "The children whom the teacher is scolding are laughing".

To account for natural languages we must therefore move up at least one level, to context-free grammars (push-down automata). The powerful new properties available are: recursiveness (a rule can accept as input the output of its previous application); abstract dependencies (a rule may apply to a whole noun phrase, or a whole clause, not just to single words); and an erasable finite-memory device (the push-down store), allowing local comparisons between one sequence and a previous one.

As it turned out, we had to move up yet another level, to context-sensitive grammars (linear, bounded automata) because we need rules sensitive to the syntactic environment

Grammar

It is useful to consider the concept of grammar as primary and that of language as derived.

(context) in which a certain string appears in a sentence. We need to modify that string, according to the rule, if and only if it is flanked in the sentence by some identifiable syntactic configuration. We have to find a class of grammars that provides a reasonable model of human linguistic ability, admitting highly constrained displacements of whole sentential constituents from one position in the sentence to another, as in "Which guest of your mother did you inadvertently insult?", with rules that systematically also apply to unexpressed linguistic components — implicit pronouns, for instance, as in "Go home!" ('you' being left implicit.) Such grammars must display the subtle difference, obvious to any English speaker, between "John is easy to please" and "John is eager to please".

These grammars are all-important to humans, but are not precisely matched by mathematical theory because they dwell at the level of a formally ill-defined, special class of context-sensitive grammars — possibly because natural languages have been shaped by the haphazard biological evolution of the human brain. Nonetheless, a new approach — the theory of generative grammars (the 'minimalist program') — is focusing on identifying the principles conceptually necessary to account for grammatical phenomena in the languages of the world, and on what excludes those phenomena that do not arise. A very narrow grammatical apparatus appears to be sufficient to account for all human grammars: concatenation (merging lexical elements one with another) and movement (the result of copying a phrasal constituent elsewhere in the sentence, concatenating it to the extant structure, then deleting one of the copies). All linguistic variation is restricted to a free combinatorial choice between two possible values assigned to each of a few parameters. Human natural grammars are trimmed to the barest essentials, with boundaries along the lines of strict conceptual necessity, rather than along the whimsical contours of evolution. ■

Massimo Piattelli-Palmarini is in the Department of Linguistics, University of Arizona, Tucson, Arizona 85721, USA.

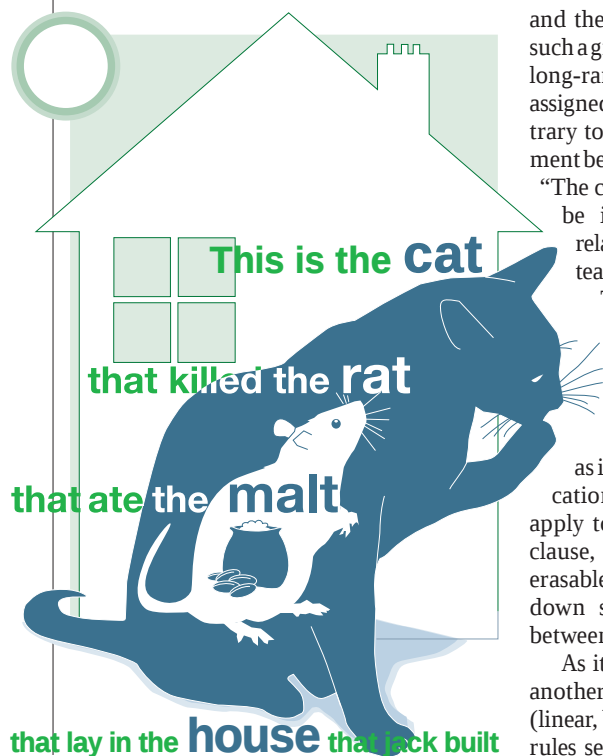
FURTHER READING

Wasow, T. in *Foundations of Cognitive Science* (ed. Posner, M.) 161–205 (MIT Press, Cambridge, Massachusetts, 1991).

Chomsky, N. *The Minimalist Program* (MIT Press, Cambridge, Massachusetts, 1995).

Pullum, G. K. & Scholtz, B. C. *Nature* **413**, 367 (2001).

VICKY ASKEW



Senseless motion

Eve Marder

Rhythmic movements such as locomotion are produced by oscillatory circuits in the central nervous system. Work in fruitflies shows that the neural circuitry for such movements develops without peripheral sensory feedback.

Human babies kick, turn and wriggle before birth. Likewise, chick, frog, fish, lobster and fly embryos move before hatching¹⁻³. In adults, rhythmic patterns of movement, such as flying, swimming, crawling and breathing, are produced by central pattern generators, groups of neurons in the central nervous system that can generate rhythmic motor patterns in the absence of input from sensory neurons at the body's periphery, although sensory input modifies the centrally generated motor patterns to the needs of the animal⁴. The precocious movements seen in embryos are thought to result from early activity in these pattern-generating neural circuits which will eventually develop to produce behaviourally relevant rhythmic movements after birth or hatching⁵.

On page 174 of this issue⁶, Suster and Bate describe how they have used genetics to construct *Drosophila* embryos and larvae that have either no peripheral sensory function or almost no peripheral sensory neurons. Nonetheless, these animals move. The results show that the basic structure of the central circuits that produce rhythmic movements can develop and function in the absence of sensory inputs.

Drosophila larvae move forwards and backwards by peristaltic (wavelike) crawling produced by waves of muscle contraction. The basic central circuitry that produces these larval movements is present in late-stage embryos, as such embryos also show peristaltic movements (Fig. 1). Suster and Bate⁶ used time-lapse videomicroscopy to monitor peristaltic movements in late embryos as an assay of the effect of genetically altering sensory input during the development of the circuitry that produces these movements.

In their first set of experiments, the authors created flies in which the gene for tetanus toxin was expressed in the peripheral sensory neurons that give the embryo information about its body position during normal peristalsis. The presence of tetanus toxin prevented the neurons from releasing neurotransmitter at the synapse, or junction, between neurons. It does this by interfering with the process by which vesicles containing the transmitter fuse with the neuronal cell membrane to release their contents. In contrast, non-toxin-containing motor neurons and interneurons (neurons that are neither sensory nor motor but are part of the central brain circuits) released transmitter



Figure 1 Embryo waves — peristaltic motion in a late-stage fruitfly embryo, just before hatching. These three photographs show movement beginning at the back of the embryo, on the right of the top image, passing towards the front as a wave of coordinated contractions.

normally. The embryos without sensory synapses generated both forward and backward, albeit abnormal, peristalsis, and hatched successfully from their egg cases.

In these first experiments, the sensory neurons were still present, although not releasing neurotransmitter from vesicles. Therefore, it was possible that the neurons were still exerting a role in shaping the construction of the motor circuits, either by acting as a substrate or scaffold for growth or by releasing signalling molecules not contained in vesicles. To address this possibility, the authors produced animals lacking the protein encoded by the *senseless* gene. The *senseless* protein is necessary for the normal development of peripheral sensory neurons, and in the absence of this gene almost all of the body's sensory neurons fail to develop⁷. Suster and Bate⁶ found that these embryos also continued to make rhythmic peristaltic movements, although they did not hatch successfully.

What can Suster and Bate's experiments tell us about the roles of sensory neurons in embryonic peristalsis? Both sets of genetic-

ly altered animals move, but neither set moves normally. Because a few sensory neurons persist in the absence of the *senseless* gene product, it is gratifying that essentially the same 'take-home messages' were obtained by two independent methods.

First, these experiments are an elegant use of genetics to make the point that embryonic and larval movements are produced by central pattern-generating circuits, revisiting early experiments in other systems in which workers surgically cut sensory nerves to show that rhythmic movements are centrally generated⁸. Second, it has long been recognized that sensory feedback has a crucial role in shaping the motor patterns produced by central pattern-generating circuits⁹. Therefore, it is not surprising that there were differences in the movements produced in normal and genetically altered animals lacking sensory function. In both sets of altered animals, forward peristalsis was less frequent than in control animals. So even at this early stage of development, sensory feedback may be helping the animal to control the direction of its movement, which is essential if the animal is to adapt its movements to the world's demands.

Suster and Bate⁶ demonstrate that sensory transmission is not required for the development of a circuit that is adequate for producing rhythmic movements. But they do not demonstrate that the circuit that forms in the absence of sensory information or most sensory neurons is identical to that which forms in their presence. Theoretical and experimental work shows that similar activity patterns can be produced by different underlying mechanisms at the level of both single neurons and neuronal networks^{10,11}. Moreover, previous work from Bate's laboratory has shown that voltage-dependent currents in *Drosophila* neurons are altered when they develop in the absence of synaptic inputs from other neurons¹².

Therefore, it seems likely that the circuit that forms in the absence of sensory input or sensory neurons is altered to compensate in some way for that absence. Because movement is essential for survival, there may be many complementary mechanisms that enable the animals to construct circuits that can generate 'good-enough' movements, even when deprived of some of their most important neuronal inputs and in the absence of some of the usual molecular signals.

MICHAEL BATE & MATTHIAS LANDGRAF



100 YEARS AGO

The appearance on clear evenings of the zodiacal light after sunset at this season of the year in this latitude is usual, and it has been frequent and beautiful to observe in this district for many nights. It would be interesting if the readers of *NATURE* could detect any definite movement of the arm of light, for much yet remains to be discovered about this phenomenon, and any observer can make this point a study... The very remarkable sunset of March 6 has probably been observed by many readers of *NATURE*. The "fire-finger" left in a perfectly perpendicular position for upwards of fifty minutes after sunset was visibly withdrawn, without losing colour or size or changing from the perpendicular, and was a vivid and beautiful adjunct to a sunset afterglow strangely reminding one of the "Krakatoa sunsets" of years ago. This finger of fire the writer has only observed once before, after a similar-coloured sunset over the estuaries of the Medway and Thames last summer, but London smoke dimmed the effect.

From *Nature* 13 March 1902.

50 YEARS AGO

A conference was held during January 19–20 in London of the organization Science for Peace, which was attended by 180 scientific workers... After lengthy discussion, a statement of principles was adopted. Recognizing the danger of a third world war the statement runs... "We think it our duty to ensure that science is used solely to improve the conditions of life and to advance the arts of peace". It is the scientist's duty "to inform and enlighten the public both about the benefits that science can bring, and about the destructiveness and misery of modern war". ... What people want to hear about is not the horrors of atomic warfare, not the A.B.C. of atomic energy, but dispassionate analysis of present deadlocks, and expert suggestions as to what can be done... In so far as we contribute to the release of tension among the nations, we contribute not only to the safety and independence of our country, but also to the support and sustenance of the great scientific tradition in which we have been fostered and to removal of the suspicion that scientists are indifferent to the often revolutionary consequences of their own discoveries.

From *Nature* 15 March 1952.

It will be fascinating to see how the neuronal properties and synaptic connectivity of the circuits that control movement in these genetically altered animals compare with those that develop normally. We can also look forward to genetic manipulations in other organisms that are better suited to the study of reflex pathways in the control of movement, so that we can discover in detail the developmental consequences of altered sensory signalling.

Eve Marder is at the Volen Center, MS 013, Brandeis University, Waltham, Massachusetts 02454, USA.

e-mail: marder@brandeis.edu

Cosmology

Maintaining the standard

George F. R. Ellis

Cosmologists continue to probe for weaknesses in the 'standard model' for explaining the structure of the Universe. Happily, the model passes the latest observational test of its consistency.

There might seem to be a paradox at the heart of cosmology. On the scales of stars, galaxies and clusters of galaxies, the structure of the Universe is complex, but cosmologists claim that at larger scales it becomes very simple. The standard model of cosmology assumes that the structure of the Universe is extremely smooth at the largest observable scales, but there are other models that could account for the uniformity we see in observations made at such scales. Fortunately, there are various observational tests that can investigate the consistency of the standard model — and the results of one such test are reported by Blake and Wall¹ on page 150 of this issue.

At large scales, such as the Hubble scale (about 10^{10} light years), which is estimated using the rate of expansion of the Universe since the Big Bang, the structure of the Universe seems to follow the model codified by Robertson and Walker^{2,3}. In this model, space is everywhere isotropic (there are no preferred directions) and homogeneous (all physical and geometrical properties are the same everywhere at a given instant). So, apart from a possible uniform curvature, the model contains no spatial structure at all; things only vary with time. It is hard to get simpler than that.

One might be excused for being a mite cynical about the claim that everything is so simple at these scales. In fact, on the largest scales (beyond those that are observable), cosmologists believe that the Universe might again be complex, having a fractal structure, that is, one that has similar structure on all scales. This is a consequence of popular models^{4,5} that describe the Universe as chaotic and inflationary — that at one stage

1. Fénelon, V. S., Casasnovas, B., Simmers, J. & Meyrand, P. *Curr. Opin. Neurobiol.* **8**, 705–709 (1998).
2. Bate, M. *Curr. Opin. Neurobiol.* **9**, 670–675 (1999).
3. O'Donovan, M. J. *Curr. Opin. Neurobiol.* **9**, 94–104 (1999).
4. Marder, E. *Nature* **410**, 755 (2001).
5. Bekoff, A., Nusbaum, M. P., Sabich, A. L. & Clifford, M. *J. Neurosci.* **7**, 2320–2330 (1987).
6. Suster, M. L. & Bate, M. *Nature* **416**, 174–178 (2002).
7. Nolo, R., Abbott, L. A. & Bellen, H. J. *Cell* **102**, 349–362 (2000).
8. Marder, E. & Bucher, D. *Curr. Biol.* **11**, R986–R996 (2001).
9. Pearson, K. G. *Curr. Opin. Neurobiol.* **5**, 786–791 (1995).
10. Goldman, M. S., Golowasch, J., Marder, E. & Abbott, L. F. *J. Neurosci.* **21**, 5229–5238 (2001).
11. Golowasch, J., Casey, M., Abbott, L. F. & Marder, E. *Neural Comput.* **11**, 1079–1096 (1999).
12. Baines, R. A., Uhler, J. P., Thompson, A., Sweeney, S. T. & Bate, M. *J. Neurosci.* **21**, 1523–1531 (2001).

in its evolution it underwent exponential expansion and that this might still be occurring in some regions today. So might cosmologists be deceiving themselves when they embrace such idealized models for the observable region of the Universe?

But there is evidence for this simplicity — it is the extraordinary large-scale isotropy of the Universe around us. The distributions of very distant galaxies and radio sources are isotropic to a high degree of accuracy, as is the cosmic microwave background (CMB). This is the radiation that is believed to have originated at the moment, 300,000 years after the Big Bang, when the Universe had cooled sufficiently for matter and radiation to become decoupled — electrons and protons began to combine to form atoms, and radiation could propagate freely.

Today, we can detect this relic of the Big Bang, but there is in fact a simple, detectable anisotropy — the temperature of the CMB is 2.7 K, but, when the whole sky is observed, the CMB seems slightly hotter in one direction and slightly cooler in the opposite direction (Fig. 1). The difference is only one part in 1,000, but it creates a dipole structure in the CMB⁶. This anisotropy can be removed by factoring in a change of velocity: the dipole is interpreted as being due to the Earth moving at a speed of about 370 km s^{-1} relative to the 'rest frame' of the CMB. Once this velocity dipole is removed, the radiation becomes astonishingly isotropic⁷ — to one part in 100,000.

Much effort is now going into measuring and analysing this tiny residual anisotropy in the CMB, for it is the signature of seed perturbations that existed at the time of the matter–radiation decoupling, from which

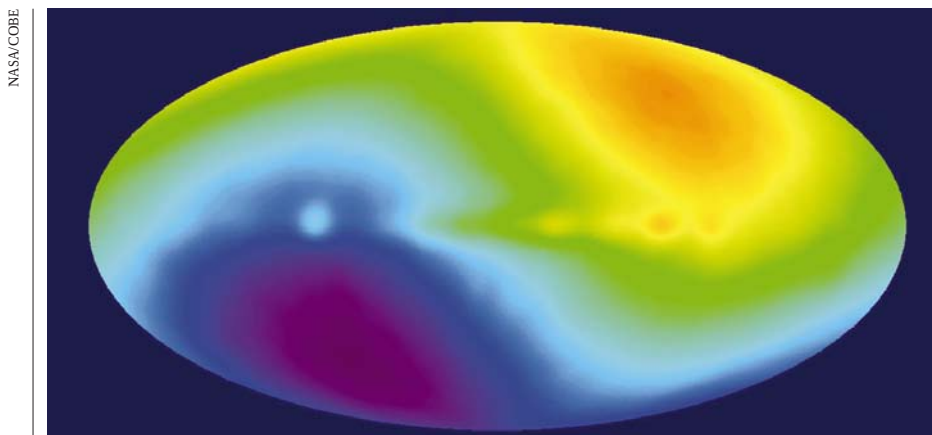


Figure 1 The all-sky map of the cosmic microwave background radiation reveals a dipole structure. Radiation in the direction of the Earth's motion appears bluer and hotter, and radiation in the other half of the sky appears redder and cooler. Blake and Wall¹ have found a corresponding dipole in the distribution of distant radio sources. This supports the belief that the microwave background is a cosmic, not a local, phenomenon and is a relic of the Big Bang.

galaxies and clusters of galaxies would eventually grow. But the point here is not what the CMB anisotropy represents, but that it is so small. The simplest explanation is that the Universe is indeed extraordinarily smooth on the largest scales — that it is a Robertson–Walker model, with tiny fluctuations superimposed.

This is not the only possible interpretation, but it is supported by a physical model. The inflationary-universe theory^{4,5,8} suggests that such a high degree of smoothness would necessarily result from a vast exponential expansion of the Universe at very early times. It also provides a mechanism for the formation of structure that predicts anisotropies in the CMB of the kind recently confirmed⁷.

But other possible explanations exist. The observed dipole anisotropy could, for example, be caused by a genuine inhomogeneity of the Universe at large scales. Or the whole picture could be wrong — the 2.7-K radiation might not be a property of the whole Universe at all, but could instead be generated locally and become isotropic in the neighbourhood of the Sun⁹.

There are two crucial tests that the CMB must pass if its interpretation as a Big Bang relic is correct. First, the radiation would have cooled as the Universe expanded; thus, for radiation that has travelled very large distances to us (equivalent to looking far back in time), we should see evidence of increasing temperature, up to a limit of 3,000 K — the temperature at the time of decoupling. This can be tested by detecting the effect of the CMB temperature on intergalactic molecules, and, so far, the CMB has passed this test with flying colours¹⁰.

The second test is that number counts of distant objects (galaxies, radio sources and quasistellar objects) must also show a dipole anisotropy that is in the same direction as that of the CMB, at about the 2% level¹¹. This

imbalance should appear because, as the observer on Earth moves towards the sources, their surface brightness will appear enhanced, so that previously undetectable sources will be raised above the detection limit. The reverse should happen as the observer moves away from sources in the other half of the sky. If this anisotropy were not found, it would spell disaster for the standard model — either the background radiation could not be of cosmological origin, or the Universe would not fit the Robertson–Walker model.

This crucial test has now been carried out by Blake and Wall¹. They have analysed data from the sky survey of distant radio sources¹² performed by the National Radio Astronomy Observatory's Very Large Array in New

Mexico, USA. The Y-shaped array of 27 radio antennas (each 25 m in diameter) achieves the resolution of a telescope several kilometres in diameter. Blake and Wall find that there is indeed a dipole, in the same direction as the temperature anisotropy and close to the expected amplitude, after experimental effects and local clustering have been taken into account.

The result is not unexpected, but it is important nevertheless. The seemingly unlikely standard model, with its simple behaviour at large scales, has passed yet another critical consistency test. It is vital that we carry out all such tests, checking every potential weakness in the standard picture. With yet another observational success behind them, theoretical cosmologists can be pleased that their basic model remains intact. ■

George F. R. Ellis is in the Department of Mathematics, University of Cape Town, Rondebosch 7700, Cape Town, South Africa.
e-mail: ellis@maths.uct.ac.za

1. Blake, C. & Wall, J. *Nature* **416**, 150–152 (2002).
2. Robertson, H. P. *Rev. Mod. Phys.* **5**, 62–90 (1933).
3. Walker, A. G. *Proc. Lond. Math. Soc.* **42**, 90–127 (1936).
4. Linde, A. D. *Particle Physics and Inflationary Cosmology* (Harwood, Chur, Switzerland, 1990).
5. Guth, A. H. Preprint astro-ph/0101507 (2001); <http://xxx.lanl.gov>
6. Smoot, G., Gorenstein, M. & Muller, R. *Phys. Rev. Lett.* **39**, 898–901 (1977).
7. Hanany, S. *et al. Astrophys. J.* **545**, L5–L9 (2000).
8. Guth, A. H. *Phys. Rev. D* **23**, 347–356 (1981).
9. Hoyle, F., Burbidge, G. & Narlikar, J. V. *A Different Approach to Cosmology* (Cambridge Univ. Press, 2000).
10. Srianand, R., Petitjean, P. & Ledoux, C. *Nature* **408**, 931–935 (2000).
11. Ellis, G. F. R. & Baldwin, J. *Mon. Not. R. Astron. Soc.* **206**, 377–381 (1984).
12. Condon, J. J. *et al. Astron. J.* **115**, 1693–1716 (1998).

Signal transduction

Molecular ticket to enter cells

Shlomo Oved and Yosef Yarden

Just as important as starting cellular signalling pathways is switching them off again. It seems that the Cbl protein has a dual function in accelerating the degradation of certain signalling molecules.

The cells that make up our bodies live in tight-knit communities and communicate with each other almost constantly. They often do so by using messenger molecules such as growth factors; when these molecules bind to receptors on the cell surface, they activate a plethora of cellular processes that set specific gene programmes in motion. The receptors must then be inactivated so that signalling can be stopped, and breakdown of the deactivation mechanisms often leads to cancer. One of the main inactivation mechanisms entails rapid clearance of the receptor from the surface — a process termed receptor-mediated endocytosis —

and degradation of messenger–receptor complexes in an acidic cellular compartment, the lysosome. Receptors may spend hours to days at the cell surface, but they are removed within seconds after messenger binding. On pages 183 and 187 of this issue, Soubeyran and colleagues¹ and Petrelli and co-workers² shed light on the rapid sequence of events that take place in this critical window of time (Fig. 1, overleaf).

A major source of information on receptor-mediated endocytosis is the equally rapid recycling of membranes at synapses, the specialized junctions between nerve cells. When nerves are stimulated, neuro-

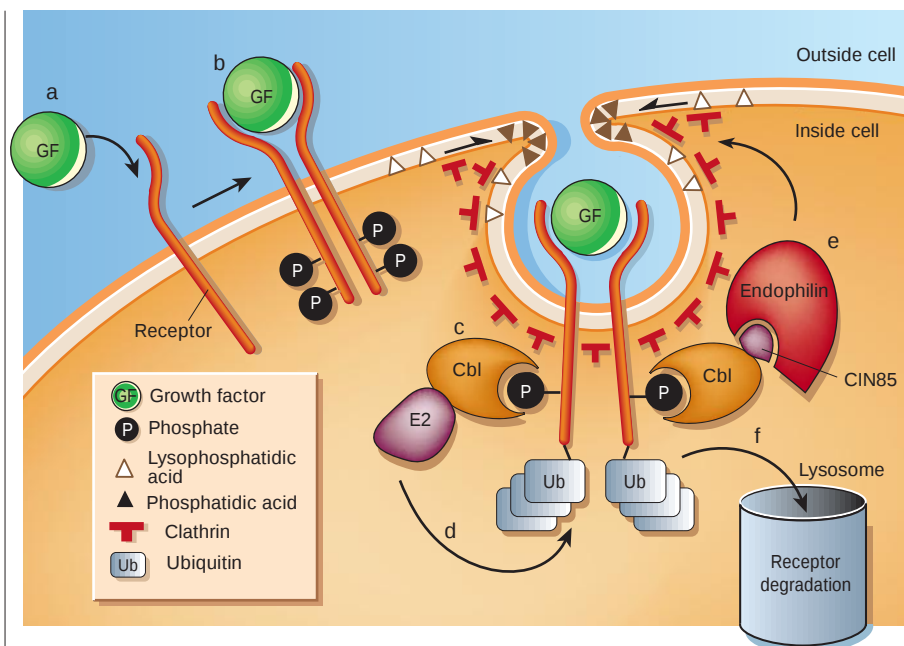


Figure 1 Switching off signalling: the dual role of the Cbl protein. a, Binding of a growth factor to its receptor is followed by b, receptor dimerization and self-phosphorylation on several tyrosine amino acids. c, Cbl is a signalling protein that anchors at specific phosphorylated tyrosines. It promotes the uptake of activated receptors through internalization of clathrin-coated membrane pits. Two activities of Cbl are needed. d, First, it attracts ubiquitin-loaded E2 molecules, which tag receptors with ubiquitin. The receptors can then be recognized and sorted into clathrin-coated areas. e, Second, Cbl attracts an endophilin–CIN85 complex, as the new papers^{1,2} show. Endophilin catalyses the conversion of lysophosphatidic acid to phosphatidic acid. This allows the inner leaflet of the plasma-membrane bilayer to adopt an extreme inward curvature, which might favour invagination and scission. f, The growth-factor–receptor complexes are transported to the lysosome and degraded.

transmitter-laden synaptic vesicles — tiny, membrane-clad sacks — fuse with the plasma membrane and discharge their contents outside the cell. As the synapse needs to stay the same size, the vesicle membrane must then be retrieved, and this is achieved by endocytosis: areas of the plasma membrane invaginate and pinch off to form intracellular vesicles. These areas can be identified because they are coated with the clathrin protein. Membrane proteins to be internalized are sorted into the invaginations, or 'pits', through adaptor proteins, which recognize specific linear amino-acid sequences in the membrane proteins.

Another important component is dynamin — a protein that hydrolyses the molecule GTP and is involved in the pinching off (scission) of the progressively invaginating pits. Although it is unclear how the GTP-hydrolysis activity regulates scission, analyses of dynamin-interacting proteins have ascribed a key function for lipid modifications. One dynamin-interacting protein, synaptojanin, removes phosphate groups from specific membrane lipids. Another, endophilin, has lysophosphatidic acid acyltransferase activity, and might be involved in altering the curvature of membranes.

Receptor-mediated endocytosis shares many similarities with vesicle recycling; for

example, clathrin-coated pits, adaptor proteins and dynamin are all involved. But, at least in the case of the epidermal growth factor receptor (EGFR), endocytosis is independent of linear sequence motifs, suggesting that membrane proteins such as growth-factor receptors are sorted into invaginating pits by a different mechanism.

The first clue to the process underlying internalization of active growth-factor receptors (also known as receptor tyrosine kinases, or RTKs) came from studies of vulva development in worms. Genetic screens identified the SLI-1 protein, a worm relative of the mammalian Cbl protein, as an inhibitor of EGFR-induced differentiation of vulva precursor cells³.

Subsequent *in vitro* studies of Cbl revealed that it helps to sort active RTKs into invaginating pits (reviewed in ref. 4; see Fig. 1). When EGFRs or other receptors bind their respective growth factors, certain tyrosine amino acids in the receptors become modified with phosphate groups (phosphorylated). The phosphorylated tyrosines bind to the so-called SH2 domain of Cbl. The region of Cbl next to the SH2 domain — a 'ring-finger' domain — then recruits an enzymatic machine that covalently tags EGFR with a 76-amino-acid molecule called ubiquitin. Tagging with a single ubiquitin molecule seems to be sufficient to

sort out those membrane proteins that are to be internalized⁵. Indeed, Cbl-mediated ubiquitination of EGFR accelerates its endocytosis and delivery to lysosomes for degradation.

So the sorting of RTKs to be internalized is controlled by two types of reversible modification — phosphorylation and ubiquitination. The recycling of synaptic proteins, by contrast, depends on sequence motifs in those proteins. But the two new papers^{1,2} suggest a key similarity between receptor-mediated endocytosis and synaptic-vesicle recycling: lipid modifications are essential for both. What's more, Cbl is involved in these modifications, as well as in protein sorting, during receptor endocytosis.

Petrelli *et al.*² used one of the non-neuronal forms of endophilin to identify a new endophilin-interacting protein — CIN85, a member of a family of adaptor proteins. The interaction occurs through a proline-rich sequence in CIN85, but the flanking SH3 domains bind to several other molecules, including Cbl. In a reciprocal approach, Soubeyran *et al.*¹ used the carboxy-terminal end of Cbl to isolate CIN85, a portion of which enabled them to unravel the interaction with endophilin. Subsequent experiments in cultured cells revealed that a complex comprising Cbl, CIN85 and endophilin is recruited to growth-factor receptors — specifically EGFR¹ and c-Met², the receptor for hepatocyte growth factor. Maximal assembly of the complex at receptors depends on phosphorylation of receptors (to recruit Cbl) and Cbl (to increase binding to CIN85).

What is the outcome of complex assembly? By using mutants of the various components and biochemical assays, the two groups^{1,2} compellingly show that, once assembled at an active RTK, the complex accelerates endocytosis, thereby terminating the biological activities of the two receptors they studied.

Perhaps most unexpected is the emerging dual function of Cbl in receptor desensitization: its ability to enhance receptor ubiquitination must be complemented by its ability to recruit endophilin molecules. How endophilin participates in receptor endocytosis remains unclear, although studies of neural synapses have provided hints. For instance, injection of anti-endophilin antibodies into the giant reticulospinal synapse of lampreys interferes with vesicle recycling⁶, and a mutant form of endophilin can inhibit the formation of synaptic-like microvesicles in an *in vitro* assay⁷. According to existing models, endophilin acts as an 'effector' of dynamin: by adding the unsaturated fatty acid arachidonate to lysophosphatidic acid (LPA, a membrane lipid), it converts this inverted-cone-shaped lipid to phosphatidic acid, a cone-shaped lipid that favours negative curvature of the plasma membrane's

inner leaflet. This is thought to facilitate the invagination of membrane pits or vesicle scission.

Rigorous proof that membrane bending by a Cbl–endophilin complex helps to dispatch ubiquitin-tagged RTKs on their long intracellular journey will require analysis by electron microscopy and precise definition of endophilin's catalytic site. Moreover, endophilin's enzymatic activity in solution is very weak (it converts just one molecule of LPA per minute); if it is important in receptor endocytosis, it has to work much more quickly in cells. Alternatively, endophilin might bend the membrane by forming a complex with other endophilin molecules and directly invaginating lipid bilayers⁸.

On the ubiquitination side, a major unanswered question is how the clathrin coat identifies ubiquitin-tagged RTKs. A ubiquitin-interacting amino-acid motif, found in both endosomal and proteasomal proteins, could provide the link, but firm evidence is lacking. More generally, because receptor sorting occurs in other endocytic compartments, as well as in the cellular pathway for protein synthesis and sorting, the

newly discovered ability^{1,2} of Cbl to couple sorting events to membrane bending may become a prototype mechanism in vesicular transport.

Shlomo Oved and Yosef Yarden are in the Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100, Israel.

e-mail: yosef.yarden@weizmann.ac.il

1. Soubeyran, P., Kowanetz, K., Szymkiewicz, I., Langdon, W. Y. & Dikic, I. *Nature* **416**, 183–187 (2002).
2. Petrelli, A. *et al.* *Nature* **416**, 187–190 (2002).
3. Yoon, C. H., Lee, J., Jongeward, G. D. & Sternberg, P. W. *Science* **269**, 1102–1105 (1995).
4. Thien, C. B. & Langdon, W. Y. *Nature Rev. Mol. Cell Biol.* **2**, 294–307 (2001).
5. Hicke, L. & Riezman, H. *Cell* **84**, 277–287 (1996).
6. Ringstad, N. *et al.* *Neuron* **24**, 143–154 (1999).
7. Schmidt, A. *et al.* *Nature* **401**, 133–141 (1999).
8. Farsad, K. *et al.* *J. Cell Biol.* **155**, 193–200 (2001).

Oceanography

An extra dimension to mixing

Chris W. Hughes

According to a new theory, stratification of the oceans is controlled by a balance between heat input at the surface and heat redistribution by eddies. But it is early days for this sea-change in thinking.

In the oceans, buoyant surface water lies above denser deep water. In most parts of the world there is a well-defined boundary between the two in which, at depths typically between 100 m and 1,000 m, the temperature changes rapidly between surface and deep values. This sharp temperature gradient is

known as the permanent thermocline. Its depth and sharpness vary around the world, and control the speed at which the ocean circulation adapts, for example, to winds, and to the currents that result from these and other 'forcing' phenomena. Understanding what sets the depth and form of the

Neurobiology

The bitter-sweet taste of amino acids

Amino acids are essential for life: they form the building blocks of the proteins that our genetic code encrypts. And thanks to nature's ingenuity, mammals have evolved a partiality for the flavour of these crucial components; some taste bitter, others sweet and some simply delicious, or 'umami' — a taste that is associated in Western countries with Chinese takeaway meals. But what are the molecular players that underlie this perceptual diversity?

In this issue (*Nature* **416**, 199–202; 2002), Greg Nelson and colleagues begin to solve the conundrum. They have characterized a mammalian amino-acid taste receptor and find that, surprisingly, this one receptor responds to several of the 20 L-amino acids that are commonly found in proteins. In another clever move by nature, the receptor is entirely unresponsive to their mirror images — the biologically less significant D-amino acids.

The amino-acid receptor is located on the surface of taste cells and is coupled to well-known molecular switches called G proteins. In fact, it consists of two different,

taste-specific G-protein-coupled receptors (GPCRs), T1R1 and T1R3. Nelson *et al.* developed a cell-based assay to identify candidate taste receptors, and found that cells expressing both T1R1 and T1R3, together with a 'promiscuous' G protein that causes the release of calcium from internal cellular stores, responded to the application of different L-amino acids with an increase in intracellular calcium concentration. Moreover, a substance that enhances the tongue's response to L-amino acids — inosine monophosphate — increased the response of the T1R1+3 receptor *in vitro* and *in vivo*.

Interestingly, when T1R3 is combined with a different taste-specific GPCR protein called T1R2, it forms a functional 'sweet receptor' (G. Nelson *et al.* *Cell* **106**, 381–390; 2000). The authors now show that this receptor is activated by a range of sweet-tasting D-amino acids. So, depending on the identity of its partner protein, T1R3 may contribute to the sensation of different tastes.

While carrying out this work, Nelson *et al.* also found a molecular explanation for the taste preferences

of different species. For example, the human T1R2+3 sweet receptor is much more responsive than its rodent equivalent to artificial sweeteners such as aspartame and cyclamate. And the human T1R1+3 amino-acid receptor is more sensitive to monosodium glutamate than to any of the amino acids. This suggests that activation of the human T1R1+3 receptor gives rise to the sensation of umami, perhaps in concert with another candidate receptor for monosodium glutamate that was described two years ago by N. Chaudhari and colleagues (*Nature Neurosci.* **3**, 113–119; 2000).

But the mystery of how different amino acids can activate just one receptor, yet give rise to a whole spectrum of taste sensations, remains to be solved. If our sense of smell is any precedent, the answer will emerge not only by working out which tastes are detected by which

receptors, but also by understanding the complex network of nerves that sends these signals to the brain.

Lesley Anson



thermocline is therefore one of the fundamental aims in oceanography.

The short answer is heating and cooling, as warmer water is less dense. But two papers by Marshall and colleagues, published in the *Journal of Physical Oceanography*^{1,2}, provide a challenge to the usual textbook answers. The authors suggest that the limiting term in the ocean budget for water buoyancy, and therefore the factor that determines the steady state for the thermocline, is the rate at which eddies are formed and transfer heat.

In one model, known as the ventilated thermocline³, wind action causes surface water to circulate to intermediate depths, resulting in vertical temperature transfer. The variation of temperature with depth, which defines the thermocline, arises as horizontal temperature variations at the surface are 'mapped' onto the vertical plane by this circulation. When the influence of these 'ventilated' waters on water that circulates without reaching the surface is taken into account⁴, the theory is quite successful in explaining many observed features of the ocean.

But the ventilated thermocline model nonetheless remains incomplete. It describes a wind-driven circulation only, and does not account for the processes that actually set up the large-scale temperature gradients in the ocean. These processes include the greater heating from the Sun in the tropics than at the poles, and the effect that this has on water movement. Certain issues of water dynamics are also swept under the carpet by assuming the existence of 'passive western boundary currents', such as the Gulf Stream, which flow near to the western edges of ocean basins. These currents are needed to 'close' the circulation, making it hard to talk about such things as heat budgets when a part of the closed circulation lies outside the theory.

An alternative to the ventilated thermocline is the 'diffusive thermocline' model⁵, in which the maintenance of large-scale temperature differences is explicitly taken into account. This is essentially a one-dimensional theory. The deep waters of the ocean are formed at only a few sites in the North Atlantic and near to Antarctica, where surface water cools, becomes denser and less buoyant, sinks to the bottom, and spreads around the world by flowing at great depth. The denser water is formed mostly by cooling, although salinity also determines density and can be important in the formation of these deep waters.

To complete the circuit, the deep water must upwell, becoming less dense as it does so. The diffusive thermocline theory assumes that downward diffusion of heat counteracts the upward movement of dense water to maintain stratification in the ocean. In subtropical regions, the wind pumps water downwards near the surface, resulting

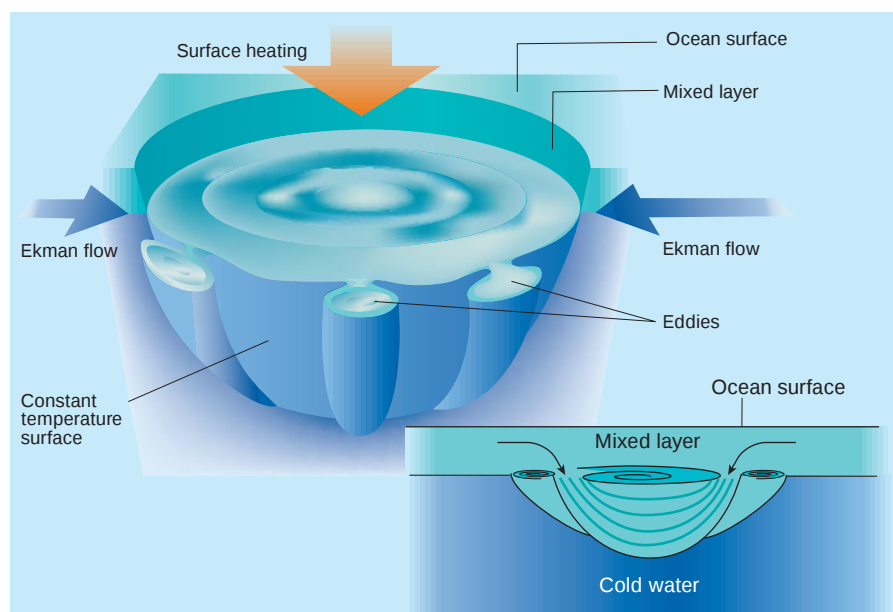


Figure 1 An oceanographic mixing bowl. The bowl shape represents a surface on which the temperature is constant, capped by an Ekman layer in which the wind directly drives water flow and mixing. The sum of the heat input due to direct surface heating and to Ekman flow must be balanced by heat transport out of the bowl's sides below the mixed layer. Marshall and colleagues^{1,2} propose that eddies breaking away from the bowl transport this heat at a rate that is determined by the bowl's shape. Applying this heat budget to a series of bowls nested inside one another (inset) results in a temperature pattern involving a region where temperature varies with depth: the thermocline.

in a confluence of shallow and deep water at a certain depth. The resulting temperature discontinuity, which is smoothed by diffusion, gives rise to the thermocline.

The principal shortcoming of the diffusive thermocline model is in reconciling the observed sharpness of the thermocline with measurements of turbulent diffusion (molecular diffusion is on a much smaller scale and can be neglected in the absence of small-scale turbulence). Measured diffusion rates⁶⁻⁸ are ten times too small to match thermocline structure, except in a few regions where the topography of the ocean floor is rough or currents are particularly strong, enhancing the turbulence.

The new theory proposed by Marshall and colleagues¹ offers a variation on the 'closed heat budget' idea of the diffusive thermocline. It extends the one-dimensional idea of vertical diffusion to three dimensions by describing how large, horizontal eddies can transport heat across a mean temperature surface. The key to this theory is to consider the heat budget of a bowl of water, bounded by a surface on which the temperature is constant (Fig. 1). In a long-term steady state, the surface heating must be balanced by a combination of diffusion and advection of heat across the bowl. Assuming the diffusion by small-scale turbulence to be negligible outside a shallow, surface mixed layer where the flow is directly driven by wind (the Ekman layer), a three-way balance results. The heat drawn into the bowl by surface heating and advection in the Ekman layer must be balanced by loss of heat in eddies breaking off

from the bowl. Although small-scale turbulence or diffusion is then required to dissipate the eddies, the factor that controls the heat budget is the rate of formation of large eddies which carry heat away from the bowl.

To estimate this rate, Marshall and colleagues assume that eddy heat flux is proportional to the temperature gradient, and to the tangential flow speed of the eddies. The tangential flow results from the pressure gradient associated with the temperature anomaly inside the bowl. The flow that arises from this radial pressure is deflected by the Coriolis force, which is due to the Earth's rotation and here creates a circulation around the bowl, so the flow speed can also be described in terms of temperature. By calculating the heat budgets of the water inside the bowl and of the surface mixed layer, a complete theory can be derived which determines the stratification that results from the imposed heating and Ekman flow — laboratory experiments show that it works. Interestingly, Ekman flow turns out to be a more important heat source for the deep ocean than direct surface heating.

The laboratory is one thing, but the real ocean is quite another. The laboratory experiments mimic a flat Earth, but the Earth's curvature and the topography of the ocean floor can both have a large impact on eddy generation, and ocean-floor topography can produce strong, narrow currents, effectively turning the neat bowl of Fig. 1 into something more like a squid with long, convoluted tentacles. Nonetheless, as Marshall and colleagues describe in their second paper², this

simple approach gives a reasonable scaling for stratification in the Southern Ocean and the resulting strength of the Antarctic Circumpolar Current (in which it is generally thought that eddy fluxes are indeed a significant factor⁹). Moreover, predictions for other ocean basins give reasonable thermocline structures, hinting that this concept has much wider applicability than in just the Southern Ocean, and that a complete thermocline theory must consider the role of eddies.

It is early days for an idea that has not yet been adapted to realistic ocean conditions. But the first signs are encouraging: it may be large, horizontal eddies — rather than small, vertical ones — that control the ocean. ■

Chris W. Hughes is at the Proudman Oceanographic Laboratory, Bidston Observatory, Prenton, Merseyside CH43 7RA, UK.
e-mail: cwh@pol.ac.uk

1. Marshall, J., Jones, H., Karsten, R. & Wardle, R. *J. Phys. Oceanogr.* **32**, 26–38 (2002).
2. Karsten, R., Jones, H. & Marshall, J. *J. Phys. Oceanogr.* **32**, 39–54 (2002).
3. Luyten, J. R., Pedlosky, J. & Stommel, H. *J. Phys. Oceanogr.* **13**, 292–309 (1983).
4. Rhines, P. B. & Young, W. R. *J. Fluid Mech.* **122**, 347–367 (1982).
5. Robinson, A. & Stommel, H. *Tellus* **11**, 295–308 (1959).
6. Ledwell, J. R. *et al. Nature* **364**, 701–703 (1993).
7. Polzin, K. L. *et al. Science* **276**, 93–96 (1997).
8. Heywood, K. J., Naveira Garabato, A. C. & Stevens, D. P. *Nature* **415**, 1011–1014 (2002).
9. Rintoul, S. R., Hughes, C. W. & Olbers, D. J. in *Ocean Circulation and Climate: Observing and Modelling the Global Ocean* (eds Siedler, G., Church, J. & Gould, J.) 271–302 (Academic, London, 2001).

Cell biology

A new view of photoreceptors

Franck Pichaud and Claude Desplan

The light-gathering structures in our eyes are specialized membranes found on cells known as photoreceptors. Two studies show that a protein called Crumbs is crucial for the development of these membranes.

The striking conservation of gene function from fruitflies to humans is under the spotlight again. On pages 143 and 178 of this issue, Pellikka and colleagues¹ and Izaddoost and co-workers² describe the role of the Crumbs protein in controlling the development of photoreceptors — the light-detecting cells in the eye — in fruitflies. The story is fascinating in itself, but also has implications for our understanding of the forms of two inherited human eye disorders, retinitis pigmentosa and Leber congenital amaurosis^{3,4}, in which CRB1, the human counterpart of Crumbs, is mutated. Indeed, much of what is already known about the CRB1 protein is based on its physical and functional similarity to the fruitfly Crumbs.

The study of developmental genes in fruitflies (*Drosophila melanogaster*) has provided deep insights into complex cellular processes that do not exist in simpler model systems such as yeast. One such process is that by which the 'top' and 'bottom' of an epithelial cell become distinct — a feature that ensures that epithelial tissues, such as those that cover the skin and line the gut, retain their highly polarized architecture. At the cellular level, this polarity relies on the vectorial segregation of different proteins to specific regions of the plasma membrane, thereby allowing a cell to distinguish top from bottom⁵.

For instance, a complex consisting of the proteins Crumbs, Discs lost and Stardust, together with another complex involving Bazooka, specifies the apical membrane (that at the top) of *Drosophila* epithelial cells^{6–9}. The basolateral membrane (that at

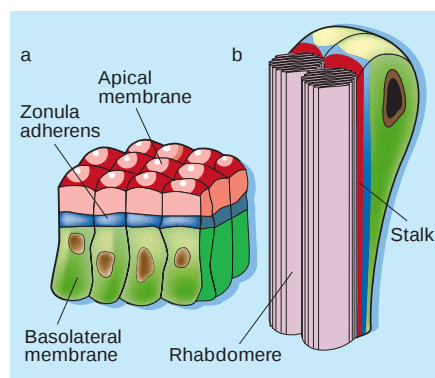


Figure 1 Determinants of cellular polarity. a, Epithelial cells in the fly embryo. The apical part of the cell is in red, the zonula adherens is blue, and the basolateral part is green. The proteins Crumbs, Stardust, Discs lost, β -spectrin, Bazooka, protein kinase C and DPA-6 are present in the red areas; Armadillo/ β -catenin, Canoe, Pyd and spectrins in the blue areas; and Lethal giant larvae, Discs large, Scribble and myosin II in the green areas. b, The same colour code is used for photoreceptors. The photoreceptor could be viewed as the equivalent of epithelial cells, but turned by 90°. The rhabdomere (light-gathering structure) is shown in purple; although the most apical membrane, it lacks Crumbs.

the bottom) is characterized by a complex comprising Lethal giant larvae, Discs large and Scribble^{10,11} (Fig. 1a). Two narrow circumferential membrane domains form at the boundary between apical and basolateral membranes; these domains are needed to

allow neighbouring epithelial cells to interact. One of the domains, the zonula adherens, is enriched in the cadherin–catenin protein complex. The other, the marginal zone, is found just apical to the zonula adherens and is enriched in apical determinants such as Crumbs.

Photoreceptors are a type of epithelial cell that undergoes a massive change in shape during development: their apical–basal axis rotates by 90° and the apical membranes elongate to form a long, densely packed array of light-sensitive membranes (Fig. 1b). These specialized membranes are called outer segments in vertebrates and rhabdomeres in flies. They originate from different apical extensions in vertebrates and invertebrates (from 'cilia' or 'microvilli', respectively¹²) and have different mechanisms for transducing light. Nevertheless, the final morphology is quite similar: both structures consist of packed apical membranes with high concentrations of a light-detecting pigment. The rhabdomere is supported by a specialized membrane, the stalk (Fig. 1b), which connects it to the zonula adherens; the vertebrate outer segment is similarly supported by the inner segment.

Knowing that Crumbs is involved in specifying apical membranes in *Drosophila* epithelial cells during development, Pellikka *et al.*¹ and Izaddoost *et al.*² wanted to find out whether this protein is also required for the massive changes that take place in the apical membranes of photoreceptors. Using the full range of tools available to *Drosophila* geneticists, including disrupting the function of the Crumbs gene in fly eyes, both groups discover a role for Crumbs in maintaining the integrity of the zonula adherens during photoreceptor development. In other epithelial cells, Crumbs is likewise involved in maintaining the zonula adherens, in order to specify the apical and basolateral membranes. But in photoreceptors this role is adapted to enable the rhabdomeres to elongate substantially. In addition, Pellikka *et al.* show that Crumbs has yet another function — to modulate the length of the stalk, recruiting new membrane as necessary to extend the stalk.

Interestingly, these two distinct processes — rhabdomere elongation and stalk extension — are mediated by two distinct regions within the Crumbs protein^{1,2}. Crumbs is a transmembrane protein containing a sequence of amino acids known as a PDZ-binding motif (a protein-interaction motif) within its short intracellular region. The intracellular portion of Crumbs seems to be the only part needed for the integrity of the zonula adherens, and to specify apical membranes and facilitate rhabdomere elongation^{1,2}. The very end of the intracellular domain recruits the Stardust protein, with the PDZ-binding motif recruiting Discs lost^{6–8}.

The extracellular part of Crumbs, on

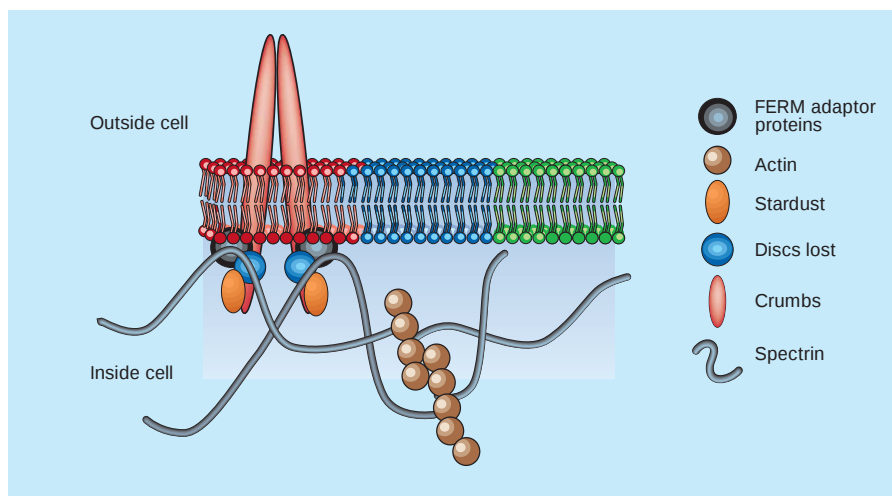


Figure 2 How the unusual architecture of the photoreceptor apical membrane might develop (based on refs 1, 2). The photoreceptor plasma membrane is divided into stalk (red), zonula adherens (blue) and basolateral parts (dark green). Here, two Crumbs proteins interact through their extracellular domains. When the complex is attached to the membrane, the regions just inside the membrane can recruit FERM proteins and spectrin, anchoring the membrane to the actin cytoskeleton. This might lead to stalk elongation by decreasing the rate of membrane internalization. The intracellular regions of Crumbs also recruit the Stardust and Discs lost proteins; together, these three proteins contribute to the other function of Crumbs — specifying the apical membrane.

the other hand, appears to be involved in modulating stalk length¹. Pellikka *et al.* found that overproduction of a truncated version of Crumbs lacking the intracellular domain could increase stalk length in a dose-dependent manner. This is particularly exciting as the extracellular domain is not needed for Crumbs to determine apical–basal polarity.

How does Crumbs recruit extra membranes to lengthen the stalk? Although they were not looking at stalk formation, Izaddoost *et al.*² might have uncovered a mechanism. They studied a region of Crumbs known as the juxtamembrane domain, which is found just inside the cell¹³. Izaddoost *et al.* show that this domain is responsible for allocating material such as the cadherin–catenin complex to the photoreceptor's zonula adherens, independently of an effect on apical–basal polarity. They also suggest that the juxtamembrane domain could be a binding site for a certain 'adaptor' protein, which in turn would recruit components of the cellular skeleton such as spectrin. Interestingly, Pellikka *et al.* show that one form of spectrin does form a complex with Crumbs and Discs lost.

So we propose the following model. As mentioned above, Pellikka *et al.* found that overproduction of the membrane-bound extracellular region of Crumbs leads to stalk extension. We suggest that Crumbs interacts with itself through this extracellular region (Fig. 2). The complex produced could then signal through the juxtamembrane domain and recruit the spectrin cytoskeleton, which would in turn decrease the rate at which membrane is internalized in the cell and so favour stalk extension¹⁴. Consistent with this model, the stalk is shorter when one

particular form of spectrin is mutated¹.

What light do these results^{1,2} shed on the forms of the human disorders retinitis pigmentosa and Leber congenital amaurosis in which CRB1 is mutated? In vertebrate photoreceptors, CRB1 is localized to the inner segment¹, as is Crumbs to the stalk in *Drosophila*, pointing to these membranes as being functionally equivalent. So it will be interesting to see whether CRB1 defines the length of the inner segment in vertebrates, as Crumbs does in fruitflies. Moreover, some patients with retinitis pigmentosa type 12 have mutations that affect only the extracellular domain of CRB1, which is intriguing given that this portion of Crumbs is only crucial in stalk extension¹. It seems that doctors are in a position to cast a fresh eye over these debilitating retinal diseases. ■

Franck Pichaud and Claude Desplan are in the Department of Biology, New York University, 1009 Main Building, 100 Washington Square East, New York, New York 10003-6688, USA.
e-mail: claudesplan@nyu.edu

- Pellikka, M. *et al.* *Nature* **416**, 143–149 (2002); online 17 February 2002 (DOI 10.1038/nature721).
- Izaddoost, S., Nam, S.-C., Bhat, M. A., Bellen, H. J. & Choi, K.-W. *Nature* **416**, 178–182 (2002); online 17 February 2002 (DOI 10.1038/nature720).
- den Hollander, A. I. *et al.* *Nature Genet.* **23**, 217–221 (1999).
- den Hollander, A. I. *et al.* *Am. J. Hum. Genet.* **69**, 198–203 (2001).
- Tepass, U. & Tanentzapf, G. *Annu. Rev. Genet.* **35**, 747–784 (2001).
- Bhat, M. A. *et al.* *Cell* **96**, 833–845 (1999).
- Hong, Y. *et al.* *Nature* **414**, 634–638 (2001).
- Bachmann, A. *et al.* *Nature* **414**, 638–643 (2001).
- Müller, H. A. & Wieschaus, E. *J. Cell Biol.* **134**, 149–163 (1996).
- Bilder, D., Li, M. & Perrimon, N. *Science* **289**, 113–116 (2000).
- Bilder, D. & Perrimon, N. *Nature* **403**, 676–680 (2000).
- Salvini-Plawen, L. V. & Mayr, E. *Evol. Biol.* **10**, 207–263 (1977).
- Klebes, A. & Knust, E. *Curr. Biol.* **27**, 76–85 (2000).
- De Matteis, M. A. & Morrow, J. S. *J. Cell Sci.* **113**, 2331–2343 (2000).

Daedalus

Electric waves

Wave power is one of the renewable sources of energy that the British government wants to develop. Yet many gates and nodding ducks, which are used to capture energy from the surface waves, have been destroyed by a furious sea; only devices with no moving parts seem to have a future. Daedalus recalls that waves are not merely a surface effect. Much of their energy is stored deep under the water, in the form of circular or elliptical closed currents that reach right down to the sea bottom (which is why waves gain height in shallow water). And because sea water contains salt, it conducts electricity. So a fully submerged, static, electrical device should be a good bet.

Electrochemists hold that a piece of metal dropped in an electrolyte gives off positive ions. They dissolve; it takes up the opposite negative charge and attracts them. The result is a double layer, the negative metal sheathed in positive ions.

So, says Daedalus, imagine two pieces of metal held apart in the water by an insulator. Ions would be released from both sides. If the ions are pushed from one side to the other by a current in the water, the arrangement should gain energy. DREADCO engineers are now designing webs of such electrical conductors to be implanted on the sea bed, out beyond the low-tide mark. Much of the technology has already been perfected by gas-drillers and other marine engineers. Each unit will damp the wave over it and generate rather a lumpy alternating current. Rectifiers will turn this into direct current; all the units will be coupled in series. Statistically the final voltage should be fairly smooth, its intensity depending on the vigour of the incoming sea. It will be led ashore directly, through delivery leads.

To the waves above, this steady loss of energy will 'feel' like an energy-absorbing sea bed. A big distributed damping installation should steal so much energy from the seas that only safe little waves will hit the shore. If so, the new system will not only provide energy, but will also reduce the need for sea walls and other marine defences. Even the wildest sea will give safe, useful energy.

Considerable cunning will be needed to find the best layout for the system. Daedalus likes the idea of close pairs of conductors aligned to cut the most likely wave ellipses; but the final design must depend on empirical insights. Ships should also welcome the system — it will damp rough weather. Anchoring would not be advisable, however; it would damage the system.

David Jones

Instant neural control of a movement signal

Hands-free operation of a cursor can be achieved by a few neurons in the motor cortex.

The activity of motor cortex (MI) neurons conveys movement intent sufficiently well to be used as a control signal to operate artificial devices^{1–3}, but until now this has called for extensive training or has been confined to a limited movement repertoire^{2,3}. Here we show how activity from a few (7–30) MI neurons can be decoded into a signal that a monkey is able to use immediately to move a computer cursor to any new position in its workspace ($14^\circ \times 14^\circ$ visual angle). Our results, which are based on recordings made by an electrode array that is suitable for human use^{4,5}, indicate that neurally based control of movement may eventually be feasible in paralysed humans.

We recorded the activity of MI neurons in three *Macaca mulatta* monkeys that had been implanted with a Utah intracortical multi-electrode array⁶. Monkeys used one hand to move a manipulandum that controlled the position of a cursor displayed on a video monitor. Using this position-feedback signal, monkeys tracked a continuously moving visual target that began at an arbitrary location and followed a pseudo-random trajectory^{7,8}.

We then used a linear filter method to test whether we could reliably reconstruct the hand trajectory from neural activity obtained in subsequent trials. Our linear filter method can be considered as a weighted sum of neural firing (here grouped in 50-ms bins) that takes into account the previous 1 s of activity^{7–9} (see supplementary information). Reconstruction accurately reflects hand trajectory, accounting for over 60% of the variance that was present in actual hand movements ($r^2 > 0.6$).

One of the monkeys then used neural control to move the feedback cursor to different targets (Fig. 1). For this closed-loop version of the experiment, reconstructions were computed continuously online in real time. An average of 1 min of continuous tracking and recording was used to construct preliminary linear filters. Once this model was constructed, the continuous task was switched to a new task that required the cursor to be moved to stationary targets (size, 0.6°), which were displayed one at a time at random locations on the monitor. During this time, hand control of the cursor position was substituted with neural control. For a period averaging 2 min, we constructed new decoding filters to relate firing to target position.

Task performance continued under two conditions: the feedback cursor could be controlled either by hand motion or by the

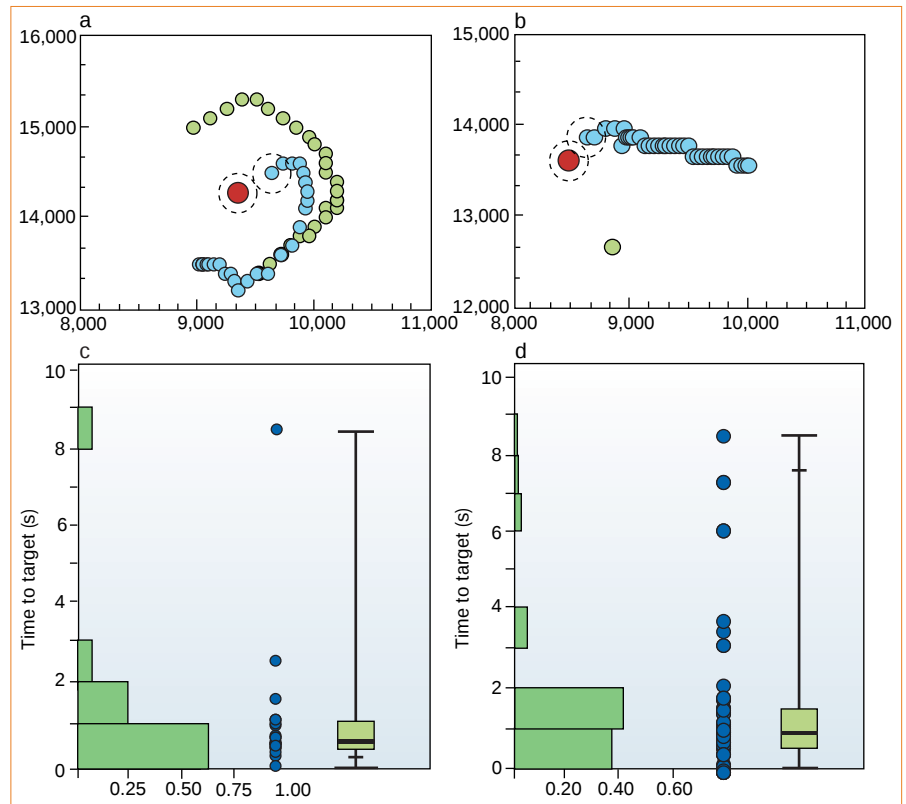


Figure 1 Performance of a target acquisition using hand or neural control. **a, b**, Trial examples showing the movement by hand (green) and by neural reconstruction (blue) of a cursor to a target (red). Dotted outlines represent the actual circumference of the target and cursor on the screen. In **a**, hand motion resembles the neurally controlled cursor path; in **b**, no manipulandum motion occurred, but the neurally controlled cursor reached the target. Each dot represents an estimate of position, updated at 50-ms intervals. Axes are in x, y screen coordinates (1,000 units corresponds to a visual angle of 3.5°); note that the two trials take place in different parts of the workspace. **c, d**, Comparable times are taken to reach the target under hand (**c**, $n = 20$) and neural (**d**, $n = 70$) control. Each panel shows three plots: data frequency distribution (histogram), trial times (spheres) and summary statistics (box and whiskers). The summary shows the median time taken to reach the target (horizontal line), the 25th and 75th percentiles (top and bottom of shaded box), and the range of the data (vertical lines).

neural signal in pseudo-randomly interleaved trials. The monkey immediately used the neural-activity-based signal to carry out the task without any further training (see movie in supplementary information). During this time, the monkey intermittently made arm movements, including, but not restricted to, target-directed hand motions (Fig. 1a). During some trials, however, the manipulandum was not moved; other muscle contractions may have occurred at this time, but these were not monitored (Fig. 1b).

Targets were considered to be acquired only if they were reached within 20 s of their appearance. We found that cursor control was nearly as good as direct hand-positional control. The time required to acquire targets using the neural signal was only slightly greater than that required for hand motions (Fig. 1c, d); this difference was not significant at $\alpha = 0.05$ by a two-

sided Kolgorov–Smirnov test. By contrast, target acquisition was unsuccessful (> 1 min) if filters were randomly shuffled between cells or if a random number was added to the filter coefficients, indicating that properly constructed filters are necessary for neural control to operate. The marked, immediate success of the linear-regression method compared with other approaches may be related to the lack of strong assumptions about neuronal firing, to the power of the linear-regression method, or to the number and type of cells used for decoding.

Neuroprosthetic devices will comprise two learning systems: a mathematical algorithm and the subject's brain. A primary role of the algorithm is to transform neural activity into a control signal that operates in a functionally usable region of space. The utility of filters constructed during neural control suggests that visual and

other feedback, coupled with a subject's dynamic learning, can compensate for inaccuracies in the model to provide an easily and voluntarily adjusted control signal. Our results demonstrate that a simple mathematical approach, coupled with a biological system, can provide effective decoding for brain-machine interfacing, which may eventually help to restore function to neurologically impaired humans.

Mijail D. Serruya, Nicholas G. Hatsopoulos*, Liam Paninski*, Matthew R. Fellows, John P. Donoghue

Department of Neuroscience, Box 1953, Brown University, Providence, Rhode Island 02912, USA
e-mail: mijail_serruya@brown.edu

*Present addresses: Department of Organismal Biology and Anatomy, University of Chicago,

Chicago, Illinois 60637, USA (N.G.H.);
Center for Neural Science, New York University,
New York, New York 10003, USA (L.P.)

1. Wessberg, J. *et al.* *Nature* **408**, 361–365 (2000).
2. Taylor, D. M. & Schwartz, A. B. *Proc. 6th Annu. Conf. Intl Funct. Elec. Stim. Soc.* 132–134 (2001).
3. Chapin, J. K., Moxon, K. A., Markowitz, R. S. & Nicolelis, M. A. *Nature Neurosci.* **2**, 664–670 (1999).
4. Maynard, E. M., Nordhausen, C. T. & Normann, R. A. *Electroenceph. Clin. Neurophysiol.* **102**, 228–239 (1997).
5. Kennedy, P. R. & Bakay, R. A. *NeuroReport* **9**, 1707–1711 (1998).
6. Maynard, E. M. *et al.* *J. Neurosci.* **19**, 8083–8093 (1999).
7. Paninski, L., Fellows, M. R., Hatsopoulos, N. G. & Donoghue, J. P. *Soc. Neurosci. Abstr.* 665.9 (1999).
8. Paninski, L., Fellows, M. R., Hatsopoulos, N. G. & Donoghue, J. P. *J. Neurosci.* (submitted).
9. Serruya, M. D., Paninski, L., Fellows, M. R., Hatsopoulos, N. G. & Donoghue, J. P. *Biol. Cybern.* (in the press).

Supplementary information accompanies this article on *Nature's* website (www.nature.com)

Competing financial interests: declared (see online version).

Sonography

Dizygotic twin survival in early pregnancy

It has been suggested that losses of twin conceptuses in very early pregnancy are high, and that for every liveborn twin pair there are a further 10–12 twin pregnancies that end up as a singleton birth¹. Here we show that in a group of women who had double-ovulated and conceived, the probability of the second egg also becoming fertilized and developing is 20–30% — which is comparable to the probability of conception and survival of a single conceptus². We conclude that the presence of one embryo does not affect the development of its twin.

So far, no direct measure has been available of the proportion of double ovulations that lead to twins. We have obtained this information by using ultrasound to identify pregnant women with two corpora lutea, and correlating these with the number who had one or two fetuses (dizygotic twins) in early pregnancy. The corpus luteum is an endocrine organ that develops in the ovary at the site at which the egg was released, and can therefore act as an indicator of the number of ovulation events.

As only two small ultrasound studies have identified the corpus luteum in early pregnancy^{3,4}, we confirmed these findings in a larger series at our centre, where the ovaries of all pregnant women are routinely examined. We scan mainly low-risk pregnancies,

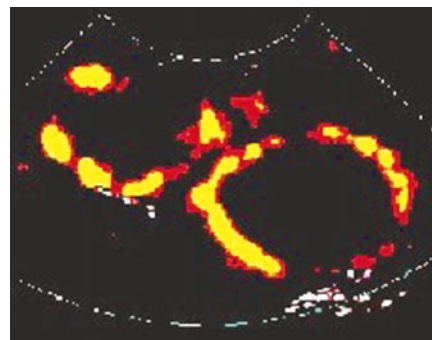


Figure 1 Doppler ultrasound of two corpora lutea in one ovary (signifying dual ovulation). The corpus luteum typically appears as an echo-filled cyst with a ring of peripheral vascularity. Colours reflect variations in blood flow, with yellow as the fastest.

as well as women who had assistance in conceiving. In scans of 504 women where both ovaries were seen in early pregnancy (5–9 weeks gestation), the corpus luteum could be identified in 94.6% of cases. Its mean diameter was 19.6 mm (± 5.28 standard deviation). Single ovulations were distributed equally between the two ovaries, occurring on the left in 49.3% of cases.

There were 48 cases of double ovulation identified by these ultrasound scans (Fig. 1). Of these (Table 1), 27 were spontaneous, with 9 among this group conceiving twin pairs (30%) and the remainder singletons. Fifteen double ovulations were induced by clomiphene citrate, and among these there were three sets of twins (20%). We were unable to determine whether the dual ovulation was spontaneous or induced in the remaining six cases, of which three were twin pregnancies. Maternal age among those who had double-ovulated (32.5 years) was not significantly different from those who had had a single ovulation (30.7 years; $P=0.07$, t -test). All sets of twins were of dichorionic and diamniotic placentation on ultrasound examination, which is consistent with dizygotic twinning.

We conclude that the presumption of

huge losses of dizygotic twins in early pregnancy¹ is unfounded, as we would then have seen many more double ovulations with a singleton-pregnancy outcome (signifying an aborted twin). The probability of the second egg also becoming fertilized seems to be similar to that of one egg becoming fertilized in a singleton pregnancy². The presence of one embryo therefore does not impede the development of its twin.

Our study does not, of course, eliminate the possibility that both twins might be lost at a higher rate than singletons. However, we do not believe that this would fit with our finding that the second egg has the same chance of developing as a singleton pregnancy once the first egg is fertilized.

The distribution of spontaneous double ovulations is consistent with a random spread of ovulation between left and right ovaries (Table 1). This suggests that the mechanism responsible for dual ovulation involves signalling from outside the ovary, rather than local intra-ovarian control⁵, as we would then have seen more double ovulations from the same ovary.

We have confirmed that the corpus luteum can be readily identified in an early-pregnancy scan, enabling us to characterize a significant number of double ovulations in the human. To our knowledge the last attempt to do this was in 1794, when William Hunter observed after 400 dissections of pregnant uteri: "When there is one child, there is only one corpus luteum; and two in the case of twins. ... In some of these cases, there were two distinct corpus lutea in one ovarium."⁶

Stephen Tong*, Simon Meagher†, Beverley Vollenhoven†**

*Department of Obstetrics and Gynaecology, Monash University, Monash Medical Centre, Clayton, Victoria 3168, Australia
e-mail: stephen.tong@med.monash.edu.au

†Monash Ultrasound for Women, Monash IVF, Epworth Hospital, Richmond, Victoria 3121, Australia

1. Boklage, C. E. in *Multiple Pregnancy. Epidemiology, Gestation and Perinatal Outcome* (eds Keith, L. G., Papiernik, E., Keith, D. M. & Luke, B.) 41–50 (Parthenon, New York, 1995).
2. Short, R. V. *Ciba Found. Symp.* **64**, 377–394 (1978).
3. Frates, M. C. *et al.* *J. Ultrasound Med.* **20**, 821–827 (2001).
4. Durfee, S. M. & Frates, M. C. *J. Clin. Ultrasound* **27**, 55–59 (1999).
5. Gillfillan, C. P. *et al.* *J. Clin. Endocrinol. Metab.* **81**, 1557–1562 (1996).
6. Short, R. V. in *The Ovary* (eds Zuckerman, L. & Weir, B. J.) 1–39 (Academic, New York, 1977).

Competing financial interests: declared none.

Table 1 Distribution of double-ovulation events

	Both left	One in each	Both right
Spontaneous	7	14	6
Iatrogenic	2	6	7
Unknown	3	3	0
Total	12	23	13

Forty-eight cases of double ovulation were identified by ultrasound scans of 504 women during early pregnancy. Of these cases, 27 were spontaneous, resulting in 9 sets of twins and 18 singletons; and 15 were induced, resulting in 3 twin pairs.

brief communications is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on *Nature's* website (www.nature.com) or on request from nature@nature.com

Crumbs, the *Drosophila* homologue of human CRB1/RP12, is essential for photoreceptor morphogenesis

Milena Pelliikka^{*†}, Guy Tanentzapf^{*†}, Madalena Pinto^{*}, Christian Smith[‡], C. Jane McGlade[‡], Donald F. Ready[§] & Ulrich Tepass^{*}

^{*} Department of Zoology, University of Toronto, Toronto, Ontario M5S 3G5, Canada

[‡] The Hospital of Sick Children, Arthur and Sonia Labatt Brain Tumor Research Center, and Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 1X8, Canada

[§] Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47906, USA

[†] These authors contributed equally to this work

The apical transmembrane protein Crumbs is a central regulator of epithelial apical–basal polarity in *Drosophila*. Loss-of-function mutations in the human homologue of Crumbs, CRB1 (RP12), cause recessive retinal dystrophies, including retinitis pigmentosa. Here we show that Crumbs and CRB1 localize to corresponding subdomains of the photoreceptor apical plasma membrane: the stalk of the *Drosophila* photoreceptor and the inner segment of mammalian photoreceptors. These subdomains support the morphogenesis and orientation of the photosensitive membrane organelles: rhabdomeres and outer segments, respectively. *Drosophila* Crumbs is required to maintain zonula adherens integrity during the rapid apical membrane expansion that builds the rhabdomere. Crumbs also regulates stalk development by stabilizing the membrane-associated spectrin cytoskeleton, a function mechanistically distinct from its role in epithelial apical–basal polarity. We propose that Crumbs is a central component of a molecular scaffold that controls zonula adherens assembly and defines the stalk as an apical membrane subdomain. Defects in such scaffolds may contribute to human CRB1-related retinal dystrophies.

Regionalization of the plasma membrane is fundamental to the physiology and development of epithelial cells^{1,2}. Circumferential junctional complexes partition the plasma membrane into distinct apical and basolateral domains, which often undergo considerable structural change during terminal differentiation to accommodate the functional specialization of epithelial tissues. Epithelial photoreceptor cells (PRCs) are a notable example, in which the apical membrane differentiates into a central photosensitive membrane and a surrounding supporting membrane^{3,4}.

The rhabdomere of PRCs of *Drosophila melanogaster* is a rod-shaped array of microvilli that contains the photopigment (Fig. 1a). The rhabdomere is surrounded by the stalk membrane that connects it to the zonula adherens (ZA), which separates apical and basolateral membranes. The architecture of PRCs is established during metamorphosis when the apical membranes of PRCs turn by 90° and undergo a dramatic distal-to-proximal expansion as the rhabdomere and stalk are formed^{4,5}. The apical membranes of mammalian PRCs show a similar subdivision into the outer segment, which contains the photopigment, and the surrounding inner segment, which connects to the ZA and the basolateral membrane (Fig. 1b).

The Crumbs (Crb) transmembrane protein is a regulator of epithelial polarity in *Drosophila* that localizes to the apical membrane and acts as an apical determinant^{6–9}. Mutations in human CRB1 (also known as RP12) cause neurodegenerative diseases of the retina, including retinitis pigmentosa and Leber congenital amaurosis^{10–12}. CRB1 and Crb exhibit a conserved domain architecture and their cytoplasmic domains show 54% identity¹³. Alignment of the cytoplasmic domains of Crb, CRB1, a *Caenorhabditis elegans* Crb homologue, and two additional putative human Crb-like proteins (CRB2 and CRB3; see Supplementary Information) reveals two conserved regions, including a carboxy-terminal PDZ-domain binding motif (ERLI), through which Crb interacts with the PDZ domain proteins Discs lost (Dlt) and Stardust (Sdt)^{14–17}. We investigated the distribution of Crb and CRB1 in PRCs and the function of Crb in PRC morphogenesis.

Crb and CRB1 localize to apical membrane subdomains

In undifferentiated cells of the retinal epithelium and in newly recruited PRCs of the third larval instar, Crb is found throughout the apical membrane but is concentrated at the marginal zone, a region of cell–cell contact adjacent to the ZA⁹ (Fig. 1c). When morphologically distinct rhabdomere and stalk domains are established, Crb becomes an exclusive component of the stalk membrane (Fig. 1e). At the same time, Crb and Armadillo (Arm, a marker for the ZA) signals are resolved, suggesting that Crb is excluded from the immediate vicinity of the ZA. This contrasts to the situation in undifferentiated epithelia (compare Fig. 1d and f). The stalk-specific localization of Crb is maintained in adult PRCs (Fig. 1g).

Antibodies raised against the cytoplasmic domain of CRB1 reveal its distribution at the apical membrane of all retinal epithelial cells, the rod and cone PRCs, and Müller's glia cells in the adult mouse retina (Fig. 1b, h–n). In rod PRCs, CRB1 is confined to the inner segment, where it concentrates in the vicinity of the ZA (also known as the outer limiting membrane; Fig. 1m). Also in cone PRCs, CRB1 is found in the inner segment. In addition, high levels of CRB1 are associated with the plasma membrane of the cone outer segments, showing a punctate distribution (Fig. 1h–k).

Control of rhabdomere shape and ZA formation by Crb

To analyse the *crb* loss-of-function phenotype in the *Drosophila* retina, we induced *crb* mutant cell clones in developing eyes by FLP-recombinase/FRT-mediated mitotic recombination^{18,19}. *crb* mutant clones were observed with normal frequency and size distribution of adult eyes⁷. Clones mutant for the null allele *crb*^{11A22} display minor irregularities in the arrangement of ommatidia and the position and number of interommatidial bristles, whereas clones mutant for the intermediate allele *crb*^{S87-2} show normal external morphology (Fig. 2b, c). Analysis of the internal structure of adult *crb*^{11A22} mutant ommatidia reveals essential roles for Crb in rhabdomere shape, ZA integrity, and stalk membrane formation. Rhabdomeres of mutant PRCs extend only 40–60% of

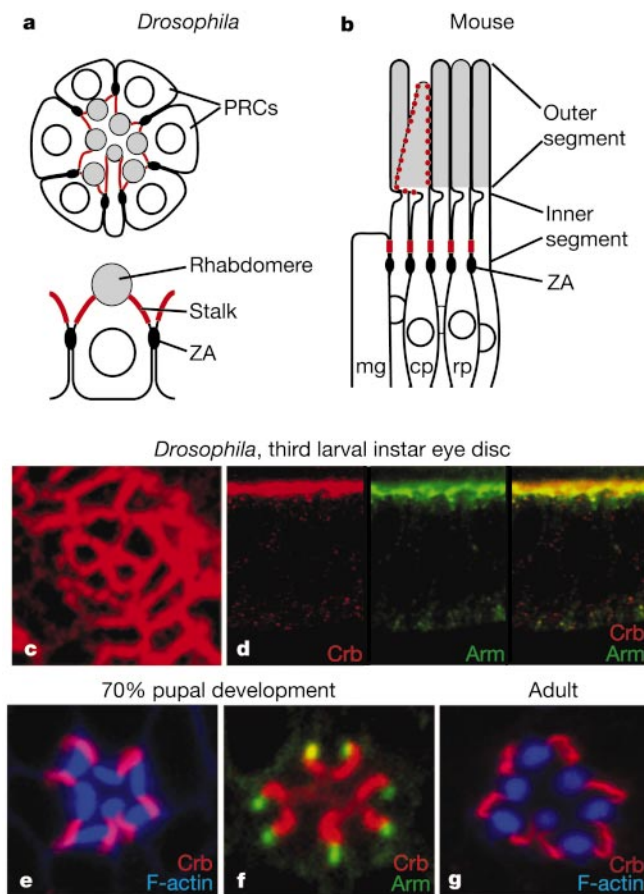
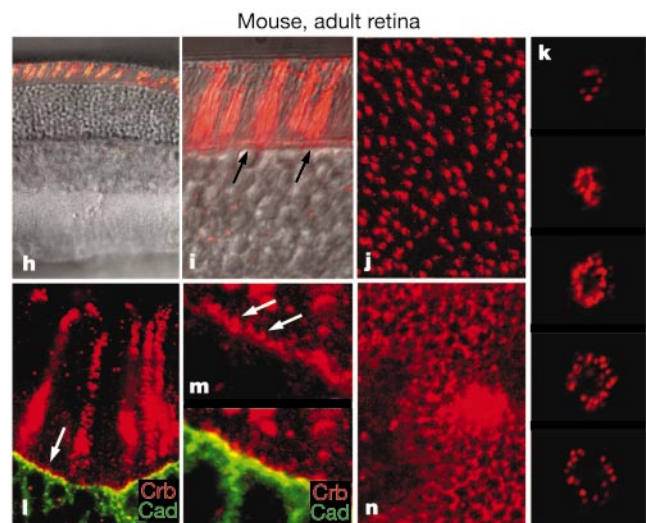


Figure 1 Distribution of Crb and CRB1 in PRCs. **a**, Schematic diagram of adult *Drosophila* PRCs forming an epithelium surrounding the inter-rhabdomeral space, into which rhabdomeres (grey) project. The apical surface of PRCs is subdivided into a central rhabdomere and a surrounding stalk membrane. Apical and basolateral surfaces of PRCs are separated by a ZA. Crb distribution is indicated in red. **b**, Schematic diagram of mouse PRCs. The ZA separates apical and basolateral membranes. The apical membrane is subdivided into inner and outer segments (grey). Crb distribution is indicated in red. cp, cone PRC; rp, rod PRC; mg, Müller's glia. **c**, Face view of the eye imaginal disc epithelium showing a cluster of developing PRCs. Crb concentrates at cell boundaries. **d**, Side view of the eye imaginal disc epithelium double labelled for Crb (red, left) and Arm (green, middle) as a marker for the ZA. Note that the signals overlap almost completely (right). **e**, **f**, Optical cross-sections through PRCs at 70% p.d. labelled for Crb (red) and F-actin (blue), which is enriched in rhabdomeres (**e**), or labelled for Crb and Arm (green, **f**). Crb is found in the

their normal distal-to-proximal length, and are confined to the distal portion of the retina (Fig. 2f). The shortened rhabdomeres are bulkier and often contact each other, which is not seen in the wild type (Fig. 2g, i). Occasionally, rhabdomeres do not face the inter-rhabdomeral space. *crb*^{S87-2} mutant PRCs display similar defects although with lower expressivity (Fig. 2e, h). In addition, secondary and tertiary pigment cells die in *crb* mutant clones (Fig. 2h, i). Pigment cells die well after eclosion when rhabdomere shape is already highly abnormal, suggesting that the death of pigment cells does not make a major contribution to the observed defects in PRC morphogenesis.

Crb has a complex role in the formation of the ZA during PRC development. *crb*^{11A22} mutant cells in third-instar larval eye discs show normal epithelial morphology and ZA integrity (Fig. 3a, b, and data not shown). The accumulation of Crb at cell–cell contacts requires Crb expression in neighbouring cells, suggesting direct or indirect interactions between Crb molecules on opposing cell membranes (Fig. 3b). A dynamic requirement for Crb in ZA formation is seen during PRC morphogenesis, when the apical



stalk membrane and is excluded from the rhabdomere. In contrast to third-instar imaginal discs (**d**), Crb and Arm signals are separated, indicating that a region 400–500 nm wide adjacent to the ZA contains little or no Crb. **g**, Optical cross-section through adult PRCs labelled for Crb (red) and F-actin (blue). **h–n**, Expression of CRB1 in adult mouse retina. CRB1 (red) is confined to the apical membranes of retinal epithelial cells (**h**, **i**). CRB1 is found in cone outer segments and in inner segments of all PRCs (arrows in **j**). **j**, Face view of a retina showing CRB1 in cone outer segments. **k**, Five optical confocal sections spaced 4 μ m apart through a cone outer segment showing punctate distribution of CRB1 in the periphery of a cone outer segment. **l**, **m**, Retinas double labelled for CRB1 (red) and cadherin (Cad; green). Arrows indicate CRB1 localization just apical to ZA (most-apical cadherin signal). **n**, Face view of retina showing reticular distribution of CRB1 in inner segments of PRCs. The intensely labelled area at mid-right represents a cone outer segment.

membrane undergoes its dramatic distal-to-proximal expansion⁵. After the initial expansion (apical membrane length about 15 μ m at 50% pupal development, p.d.) ZA are strongly fragmented and adherens junction material is scattered more widely into the lateral membrane (Fig. 3c, d). Less-severe defects are observed at 70% p.d. (apical membrane length of about 35 μ m) when most ZA are intact, suggesting a substantial recovery from early defects despite the continuous absence of Crb (Fig. 3e, f). In adult PRCs, after the final rapid elongation of the apical membrane (apical membrane length of ~95 μ m), we found minor ZA defects in the distal part of the retina, whereas ZAs were strongly fragmented in the proximal part of the retina. In the distal retina, where rhabdomeres are present, 85% of the cell–cell contacts between *crb*^{11A22} mutant PRCs display a ZA, compared with 100% in the wild type (Fig. 4c). In the proximal retina, in which rhabdomeres are absent, PRCs were difficult to identify histologically, and only scattered adherens junction material was seen. Together, these findings indicate that Crb is required to maintain integrity of the rapidly expanding ZA in an early and a late phase during PRC

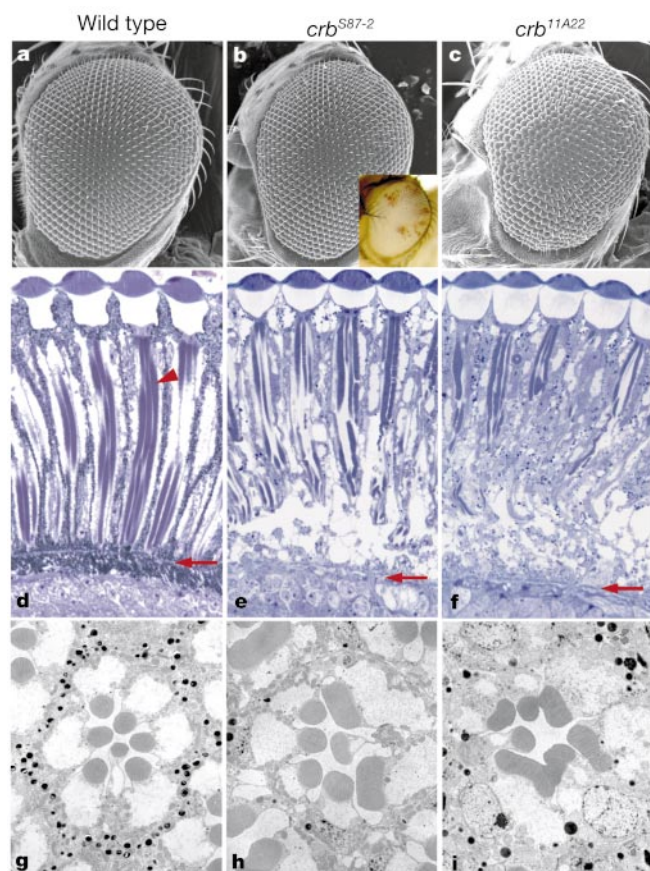


Figure 2 Adult *crb* mutant eye phenotype. **a–c**, Scanning electron micrographs of wild-type and *crb* mutant eyes. When eye clones are induced with the technique described in ref. 19, eyes contain 90% or more mutant tissue, as illustrated by the amount of white mutant ommatidia in the inset in **b**. *crb^{S87-2}* mutant eyes appear normal (**b**) and *crb^{11A22}* mutant eyes show some irregularities in the arrangement of ommatidia and a reduced number of interommatidial bristles (**c**). **d–f**, Longitudinal sections through adult retinas of wild-type and *crb* mutant eyes. Rhabdomeres (arrowhead in **d**) extend from the pseudocone to the floor of the retina (arrow). Rhabdomeres have detached from the retinal floor in *crb* mutant eyes and are shortened to about 70% (*crb^{S87-2}*; **e**) or about 50% (*crb^{11A22}*; **f**) of their normal distal-to-proximal length. **g–i**, Cross-sections of wild-type and *crb* mutant ommatidia. In the wild type, the round rhabdomeres show a stereotypic arrangement, and PRCs are surrounded by secondary and tertiary pigment cells that contain pigment granules (**g**). The cross-section profiles of rhabdomeres in *crb* mutant PRCs are enlarged, and neighbouring rhabdomeres are in contact (**h, i**). The secondary and tertiary pigment cells are missing and are replaced by apoptotic bodies. *crb* mutant cells are *white⁻* and do therefore not contain any pigment granules.

morphogenesis. Although PRCs can largely recover from early ZA defects, the pronounced ZA fragmentation seen in the proximal part of late PRCs probably contributes to the observed defects in rhabdomere shape.

Regulation of stalk membrane formation by Crb

We concentrated further analysis on the requirement of Crb in stalk formation because this function of Crb is distinct from its role in other epithelial cells. The stalk membrane of *crb^{11A22}* mutant PRCs was about 50% shorter than in the wild type (Fig. 4). In *crb^{S87-2}* mutant PRCs, the stalk was shortened by about 25%, and even in *crb^{11A22}/+* animals stalk length was on average about 10% shorter than in the wild type (Fig. 4c), suggesting a strong dose dependency for Crb in stalk formation. To determine whether Crb is sufficient to specify the stalk membrane of PRCs, we overexpressed Crb using the Gal4/UAS system²⁰. The *Rh1–Gal4* driver targets Crb overexpression to PRCs R1–R6 during and after stalk formation^{21,22}. Increased

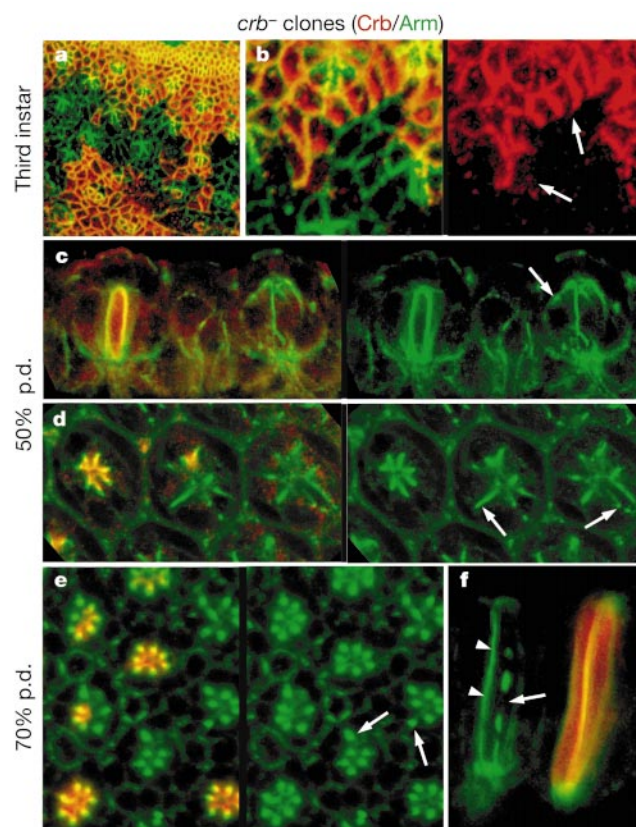


Figure 3 Development of ZA defects in *crb* mutant PRCs. Absence of red labelling indicates lack of Crb. **a, b**, Face view of third-instar larval eye—antennal imaginal discs containing *crb^{11A22}* mutant clones showing normal distribution of Arm (green) at ZA. Close-up view of the *crb* mutant clone in **a** shows that Crb accumulation at cell boundaries is not seen at contact sites between Crb-positive and -negative cells (arrows in **b**, right). **c, d**, Side view (**c**) and face view (**d**) of retina at 50% p.d. ZA are fragmented and adherens junction material is found more widely distributed in the lateral membrane in *crb* mutant PRCs (arrows). **e, f**, Face view (**e**) and side view (**f**) of *crb* mutant PRCs at 70% p.d. Arm (green) labelling indicates largely normal positioning and integrity of ZA (arrowheads in **f**) with only some ZA being slightly mispositioned or fragmented (arrows).

Crb immunoreactivity was first apparent at 75% p.d. in *Rh1–Gal4, UAS–crb* flies, and high levels of overexpression were seen at 90% p.d. with most Crb found at the apical cell pole (Fig. 5a). In many epithelial tissues, including the larval eye imaginal disc, overexpression of Crb caused mislocalization of apical markers and loss of epithelial integrity^{8,23} (data not shown), but in PRCs that overexpress Crb, apical markers such as Crb itself or Dlt remained associated with the stalk membrane and epithelial integrity was maintained (Fig. 5a). PRCs responded to Crb overexpression with a marked dose-dependent increase in the size of the stalk membrane. The additional stalk membrane was seen immediately after eclosion and persisted for at least two weeks. It formed either a spongy, convoluted mass associated with the normal stalk membrane (Fig. 5b–d), or a large flat membrane thereby expanding the inter-rhabdomeral space (Figs 4c and 5e). We did not observe a disruption of apical–basal polarity of PRCs that overexpress Crb at late pupal and adult stages. The inter-rhabdomeral space was always surrounded by a monolayer of PRCs connected by unfragmented ZA, even in ommatidia that exhibited a greatly enlarged inter-rhabdomeral space (Figs 4c and 5e, f). These findings suggest that Crb activity is sufficient to induce stalk membrane formation.

To determine whether the extracellular or the cytoplasmic domain of Crb is critical for the formation of the stalk, we expressed three truncated Crb proteins: one in which the cytoplasmic domain

is replaced by green fluorescent protein (Crb^{extraTM}-GFP), one that lacks the extracellular part of the protein except for the signal peptide (Crb^{intra}), and one that lacks the transmembrane and cytoplasmic domains and thus represents a secreted form of Crb (Crb^{extra}). Crb^{extraTM}-GFP and Crb^{extra} are effectively targeted to the stalk membrane, similar to Crb. Although Crb^{intra} is effectively incorporated into the apical membrane of undifferentiated epithelia, in PRCs Crb^{intra} accumulates in the rhabdomere (see Supplementary Information), suggesting that the extracellular domain is essential for targeting Crb to the stalk membrane. Overexpression of Crb^{extraTM}-GFP, but not of Crb^{extra} or Crb^{intra}, elicited an expansion of stalk membranes similar to that seen with full-length Crb (Figs 4c and 5g, h). In fact, expression of Crb^{extra} may have a dominant negative effect because it shortens the stalk membranes (Fig. 4c). These results indicate that Crb must be bound to the membrane to promote stalk formation, and that the extracellular domain of Crb is central in this process.

Stabilization of the spectrin cytoskeleton by Crb

To explore the role of Crb activity at the stalk membrane, we analysed the distribution of known apical proteins in *crb* mutant PRCs. β_{Heavy} -Spectrin (β_{H} -Spectrin) is associated with the stalk membrane and the terminal web that supports the rhabdomere^{24,25}. Most β_{H} -Spectrin is lost from the apical membrane of *crb* mutant

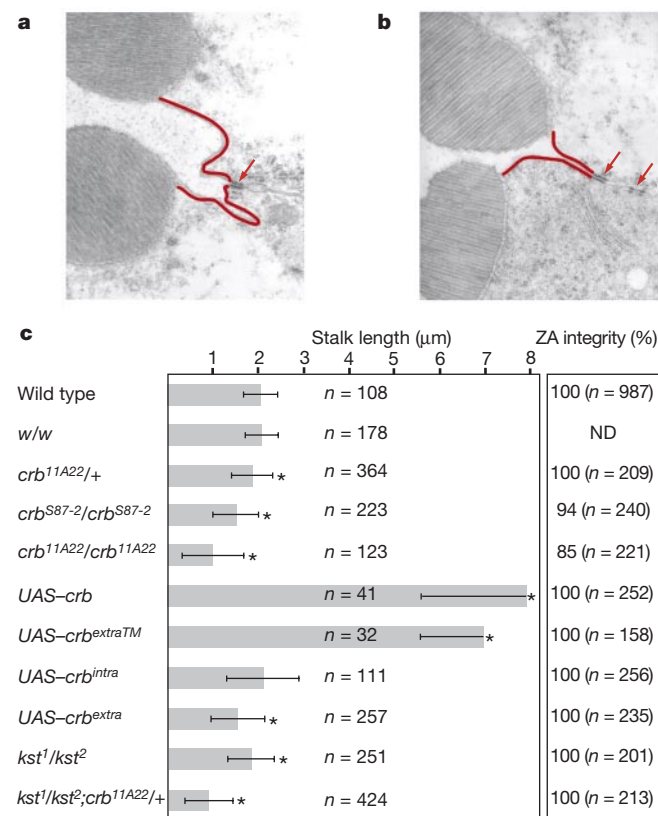


Figure 4 Quantification of stalk length and ZA integrity in *crb* and *kst* mutant eyes. **a, b**, Red lines indicate stalk membrane profiles of wild-type (**a**) and *crb*^{11A22} mutant (**b**) PRCs, which contain shortened stalk membranes. Whereas in the wild type only one adherens junction is seen between PRCs (arrow in **a**), *crb* mutant PRCs sometimes show additional adherens junctions (arrows in **b**). **c**, Stalk length was measured by tracing the stalk membrane in cross-sections of adult PRCs. Measurements of *white* (*w*) mutant eyes were included as several of the experimental PRCs were also mutant for *w*. Student's *t*-tests were performed to determine whether the observed differences in stalk length between wild-type and experimental genotypes are significant (mean $\pm$ s.e.m.; asterisk, $P < 0.0001$). ZA integrity is given as the percentage of the cell-cell contact sites between PRCs that exhibit an adherens junction. ND, not determined.

Overexpression of Crb

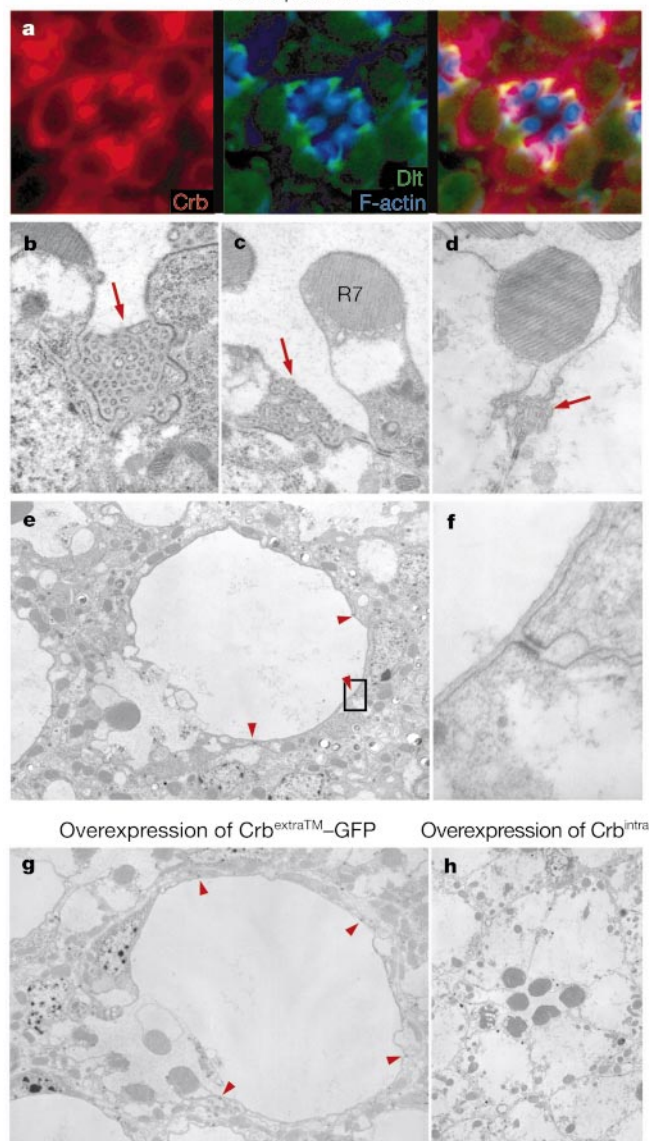


Figure 5 Crb overexpression increases the length of the stalk membrane. **a**, *Rh1-Gal4 UAS-crb* flies show high levels of Crb (red) overexpression in R1-R6 at 90% p.d. *Rh1-Gal4* is not expressed in R7 and R8, which served as internal controls. Crb (red, left) and Dlt (green, middle) are concentrated at the stalk membrane and excluded from the rhabdomere (blue, middle; labelled with phalloidin). Some Crb shows a perinuclear distribution and is presumably located in the endoplasmic reticulum. Right, merged image. **b-f**, *Rh1-Gal4 UAS-crb* adult flies raised at 29 °C exhibit large amounts of additional stalk membrane that forms either a spongiform material associated with the stalks of R1-R6 but not R7 (arrow in **b** and **c**), or an increase of the flat stalk membrane leading to an expansion of the inter-rhabdomeral space (**e**). Reduction of Crb overexpression levels achieved through the cold sensitivity of *Gal4* (flies raised at 21 °C) causes smaller amounts of additional stalk membrane to form (arrow in **d**). Arrowheads in **e** point to ZA at cell-cell contact sites between PRCs. A close-up view of the area framed in **e** is shown (**f**). Note that PRCs are connected by a ZA and that the stalk membrane is associated with a prominent extracellular lamina, whereas most of the additional luminal space is free of extracellular matrix material. **g, h**, *Rh1-Gal4 UAS-crb*^{extraTM}-GFP flies exhibit a similar increase in stalk membrane as *Rh1-Gal4 UAS-crb* flies (**g**, arrowheads point to cell-cell contacts). In contrast, *Rh1-Gal4 UAS-crb*^{intra} flies do not show an increase in the size of the stalk membrane (**h**). A second consequence of overexpression of Crb, Crb^{intra}, Crb^{extra} (not shown) and Crb^{extraTM}-GFP is a slow, progressive degeneration of the rhabdomeres (**e, g, h**). Rhabdomere degeneration may be a nonspecific effect of overexpression in this system because it is also seen when a *UAS-CDB-GFP* control construct is expressed in PRCs (not shown).

PRCs at 70% p.d. and is then located at the lateral membrane (Fig. 6a). The residual apical β_H -Spectrin is presumably located in the terminal web. In contrast to β_H -Spectrin, the concentration of α -Spectrin, which is found at low concentrations along the apical membrane, is not noticeably affected (data not shown). β_H -Spectrin is not lost from apical membranes of *crb* mutant cells in eye imaginal discs of third-instar larvae, suggesting that the dependency of apical β_H -Spectrin localization on Crb is acquired during PRC morphogenesis. In addition, the concentration of Dlt, a Crb-binding protein^{14,15}, is reduced at the stalk membrane of *crb* mutant PRCs (data not shown). These results suggest that Crb and Dlt form a complex with β_H -Spectrin. Indeed, co-immunoprecipitation

experiments demonstrated that Crb, Dlt and β_H -Spectrin are part of a complex (Fig. 6b). These findings suggest that Crb promotes stalk membrane formation by linking β_H -Spectrin to the stalk membrane.

To investigate the possibility that β_H -Spectrin operates in the Crb-dependent formation of the stalk, we analysed flies carrying loss-of-function mutations in *karst* (*kst*), the gene that encodes β_H -Spectrin, and *kst crb* double mutants. *kst* mutations are partially viable and exhibit variable defects in eye structure²⁴. Defects in rhabdome shape of *kst* mutant PRCs are less severe than those seen in *crb* mutants, and ZA are not fragmented (Figs 4c and 6d). The stalk membranes in *kst* mutants are about 10% shorter than in the wild type, and the deep folds, which are characteristic of wild-type stalks, are absent in *kst* mutants (Figs 4c and 6d). *kst* mutants carrying one copy of a *crb* null allele have normal ZA but stalk membranes are much shorter than in wild-type, *kst* homozygous or *crb* heterozygous PRCs (Figs 4c and 6e). These findings suggest that β_H -Spectrin controls size and topology of the stalk membrane and cooperates with Crb in stalk membrane formation.

Discussion

Our analysis identified two roles of Crb in PRC morphogenesis. First, Crb is required to maintain the integrity of the ZA during its rapid 100-fold expansion in PRCs. In early embryos, Crb is needed for the final step of ZA formation when patches of adherens junction material fuse into a continuous band^{9,23}. Crb probably functions similarly in embryonic ZA formation and during the expansion of the ZA in PRCs. We speculate that Crb establishes a mechanism at the transition zone of apical and lateral membranes that anchors adherens junction material and thereby supports ZA formation. ZA integrity in *crb* mutant embryos and PRCs shows recovery at later stages, and *crb* mutant cells of the imaginal disc display a normal ZA, suggesting that other mechanisms, which functionally overlap with Crb, contribute to ZA formation and maintenance. The Bazooka complex — composed of Bazooka (the *Drosophila* homologue of *C. elegans* Par-3 and mammalian ASIP), *Drosophila* Par-6 and atypical protein kinase 3 — is a candidate for such a role because it co-localizes with Crb at the apical membrane and is required for ZA formation in early embryos^{26–28}. Defects in ZA integrity could contribute to retinal dystrophy in patients carrying *CRB1* mutations. Functional redundancies among the mechanisms that stabilize cellular architecture may explain the slow progression of human *CRB1*-associated retinal degeneration^{10–12,29–31}.

Second, Crb is a central regulator of stalk membrane biogenesis. Several arguments suggest that this role of Crb is independent from its role in ZA formation. (1) Within the stalk membrane, Crb does not localize immediately adjacent to the ZA as in undifferentiated epithelial cells. (2) Overexpression of Crb enlarges the stalk membrane of PRCs but does not transform basolateral membrane to apical membrane or disrupt the ZA as seen with overexpression of Crb in undifferentiated epithelia⁸. (3) The gain-of-function effect that results from overexpression of Crb can be reproduced by expression of Crb^{extraTM}–GFP but not by expression of Crb^{intra}. In contrast, overexpression of Crb and Crb^{intra} but not of Crb^{extraTM}–GFP in undifferentiated epithelia such as the larval eye imaginal disc compromises apical–basal polarity and ZA integrity (refs 8 and 23; and G.T. and U.T., unpublished data). (4) *kst crb* double mutants have dramatically reduced stalk membranes but normal ZAs. These findings suggest that in supporting stalk membrane formation, Crb has adopted a role in PRC morphogenesis that is mechanistically distinct from its role in epithelial polarity and ZA formation. These results also reveal for the first time a mechanism that subdivides the apical membrane of epithelial cells into two different domains that support distinct structures and physiologies.

The membrane-associated spectrin cytoskeleton can determine

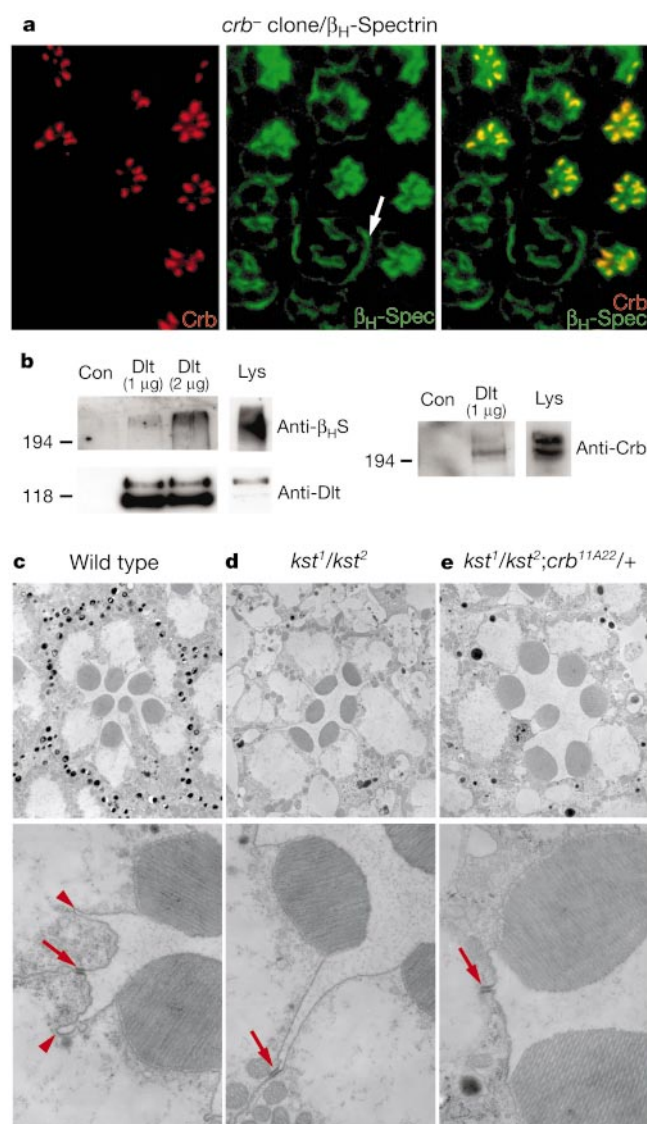


Figure 6 Crb interacts with β_H -Spectrin. **a**, *crb*^{11A22} mutant PRCs in retinas at 70% p.d. show strongly reduced levels of apical β_H -Spectrin (green) and ectopic accumulation of β_H -Spectrin at the basolateral membrane (arrow). **b**, Co-immunoprecipitation with 1 or 2 μ g of Dlt affinity-purified rabbit polyclonal antiserum precipitates β_H -Spectrin (β_H -S) and Crb from whole-cell lysates (Lys) derived from adult flies. Con, control antibody. **c–e**, Cross-sections of wild-type, *kst* and *kst crb* mutant ommatidia (close-up views shown at bottom). Arrows indicate ZA. **c**, The stalk membrane often displays deep folds (arrowheads) in the wild type. **d**, *kst* mutant PRCs display oval-shaped rhabdomeres and stalk membranes lacking deep folds. **e**, Stalk membranes are much shorter and rhabdome cross-section profiles are larger in *kst* mutant PRCs that express only one functional copy of the *crb* gene.

membrane topology and stability^{32,33}. Our findings demonstrate such a role for the spectrin cytoskeleton of PRCs, because the lack of β_H -Spectrin causes shortening and reduced folding of the stalk membrane. Loss of β_H -Spectrin may destabilize the stalk membrane causing rapid endovesiculation, as is seen when the spectrin cytoskeleton of red blood cells is compromised³⁴. Recent work has shown that coated pit budding is accompanied by a local loss of spectrin, and blocking spectrin remodelling results in the retention of coated pits at the cell surface^{35,36}. Coated pits often associate with the folds within the stalk membrane of PRCs, indicating that these folds are the site of frequent endocytosis (Fig. 6c). We hypothesize that a rate increase of endocytosis in response to lack of β_H -Spectrin could remove the membrane folds and thus straighten and shorten the stalk membrane. Conversely, stabilization of the apical spectrin cytoskeleton by Crb overexpression may prevent budding of coated pits and thus enlarge the stalk membrane. Crb must also interact with factors other than β_H -Spectrin to support stalk formation, because the loss of Crb leads to a much greater reduction in stalk size than the loss of β_H -Spectrin, and because the *kst* null mutant phenotype is dominantly enhanced by a *crb* mutation. Moreover, because the extracellular domain of Crb promotes stalk formation independently of the cytoplasmic domain, Crb is likely to interact with other integral membrane or extracellular factors that in turn can stabilize the stalk membrane. We propose therefore that Crb and β_H -Spectrin are components of an extensive molecular scaffold that integrates extracellular, integral membrane and cytoplasmic factors, and specifies topology and size of the stalk membrane.

Our analysis reveals a function for the extracellular part of the Crb protein. The dominant negative effect of Crb^{extra} expression on stalk formation suggests that the inter-rhabdomeral space contains an important ligand that interacts with the extracellular part of Crb and is titrated out by Crb^{extra}. Although Crb needs to be anchored in the stalk membrane to support stalk formation, cytoplasmic interactions are apparently not essential for this activity, as demonstrated by the effect of Crb^{extraTM}-GFP expression. The critical activity that supports stalk formation resides in the extracellular part of Crb that includes thirty EGF and four LG domains, which represent protein and carbohydrate interaction domains^{37,38}. Also the extracellular domain of human CRB1 may mediate critical interactions, because many mutations that give rise to retinal dystrophies are missense mutations that affect different EGF or LG domains of CRB1 (refs 10–12). The localization of CRB1 in the inner segment supports the notion that the *Drosophila* stalk and the inner segment of vertebrate PRCs are homologous plasma-membrane domains that rely on similar mechanisms to support their structural integrity. The localization of CRB1 in the periphery of the cone outer segments is reminiscent of the localization of the tetraspanins peripherin/RDS and ROM1 in rod and cone PRCs^{39,40}. Like these tetraspanins, CRB1 may contribute to the maintenance of cone outer segments. A role for CRB1 in colour vision may have been masked by its function in general retinal maintenance, which is apparent in CRB1-related pathologies. Further analysis of Crb and CRB1 and their interacting factors will shed light on the mechanisms that organize epithelial surface domains, and should help to understand the mechanisms of retinal differentiation and maintenance in humans. □

Methods

Fly strains and clonal analysis

The *crb* alleles we used were the null mutation *crb*^{1A22} and the intermediate allele *crb*^{S87-2} (ref. 7). *kst* alleles used were the strong/null mutations *kst*¹ and *kst*² (ref. 24). Homozygous *crb* mutant clones were generated by mitotic recombination facilitated by the FRT/FLP system. FLP expression was controlled either by a heat shock promoter (*hsFLP*)¹⁸ or the *eyelless* promoter (*eyFLP*)¹⁹. To induce clones with *hsFLP*, we heat shocked for 2 h at 37 °C during the first and second larval instar. The genotypes of flies in which clones were induced were *y w hsFLP1; FRT82B crb*[−]/FRT82B + or *y w eyFLP2 glass-lacZ; FRT82B crb*[−]/FRT82B w⁺ *cl3R3*.

Overexpression analysis

To induce overexpression of *crb*-related constructs in PRCs R1–R6, we used the Gal4/UAS system²⁰ by combining the *Rh1-Gal4* driver²¹ with *UAS-crb*^{w22}, *UAS-crb*^{extra2b}, *UAS-crb*^{extra} (ref. 8), and *UAS-crb*^{extraTM}-GFP. Flies were maintained at 29 °C to maximize Gal4 activity except for the experiment shown in Fig. 5d, in which flies were grown at 21 °C to reduce Gal4 activity thereby lowering transgene expression. *UAS-crb*^{extraTM}-GFP was constructed by first replacing nucleotides 6,558–6,650 of the *crb* complementary DNA coding for the C-terminal 31 amino acids of Crb with a linker sequence (5'-GTTGGGGG CGCTAGCGGAGCAGCGGCCGCTGGAGCTCGAGGGTGATAGGGAGCTT-3') that contains restriction sites for *NheI*, *NotI* and *XhoI*. The EGFP coding region from pEGFP-N3 (Clontech) was inserted in frame into the *NheI* and *NotI* sites of the truncated *crb* cDNA. *crb*^{extraTM}-GFP was cloned into the pUAST vector²⁰ and transgenic flies were generated using standard procedures.

Antibody production and staining

Antibodies to the extracellular domain of Crb were raised in rats. Antibodies to the cytoplasmic domain of CRB1 were raised in rabbits against the peptide ATQGTYSRQKEKE and were purified following standard procedures. For immunostaining, the following primary antibodies were used: rat polyclonal antibody (pAb) anti-Crb (F1, 1:500); rabbit pAb anti-CRB1 (851, 1:50); mouse monoclonal antibody (mAb) anti-pan-cadherin (CH-19, 1:500; Sigma) mouse mAb anti-Arm (N2-7A1, 1:50; Developmental Studies Hybridoma Bank); rabbit pAb anti-Dlt (1:200)⁴¹; rabbit pAb anti- α -Spectrin (354, 1:1000)⁴²; rabbit pAb anti- β_H -Spectrin (243, 1:250)⁴³. Secondary antibodies conjugated with Cy3, Cy5 (Jackson Laboratories) or Alexa Fluor 488 (Molecular Probes) were used at a dilution of 1:400. Rhabdomeres were stained with Alexa Fluor 488-phalloidin (Molecular Probes) at a dilution of 1:25 added together with the secondary antibodies. Pupal and adult retinas were dissected in Ringer's solution, and the cornea was removed with tungsten needles. Eye imaginal discs were dissected in PBS on ice. *Drosophila* tissue was fixed for 15 min in phosphate buffer (pH 7.2) containing 4% formaldehyde. Mouse eyes were incubated for 1 h in PBS containing 4% formaldehyde on ice. Dissected retinas were post-fixed for 20 min in 4% formaldehyde in phosphate buffer. Antibody staining followed standard procedures. Confocal images were obtained with a Zeiss LSM510 scanning laser confocal microscope using Plan-Neofluar 40X/1.30 oil and Plan-Apochromat 100X/1.40 oil lenses.

Co-immunoprecipitation experiments

Extracts were prepared from adult flies homogenized in phospholipase C (PLC) lysis buffer (50 mM Hepes at pH 7.5, 150 mM NaCl, 10% (v/v) glycerol, 1% (v/v) Triton X-100, 1.5 mM MgCl₂, 1 mM EDTA, 10 mM sodium pyrophosphate and 100 mM NaF) containing Complete protease inhibitor (Roche). The lysate was clarified by centrifugation and protein concentrations were determined by Micro BCA analysis (Pierce). We incubated 1.5 mg of total protein lysate at 4 °C for 2 h with 1 or 2 μ g of affinity-purified rabbit anti-mouse immunoglobulin control or rabbit polyclonal Dlt antiserum. The immune complexes were collected by centrifugation, washed five times in PLC lysis buffer, dissociated by boiling in Laemmli sample buffer and analysed by SDS-PAGE and immunoblotting.

Electron microscopy, ZA and stalk length quantification

For transmission electron microscopy, adult flies were immersed in ice-cold fixative (2.5% glutaraldehyde, 2% formaldehyde, 0.1 M sodium cacodylate, pH 7.4) and heads were severed and bisected. After fixation at 4 °C for 3 d, eyes were treated with 1% osmium tetroxide in 0.1 M sodium cacodylate (pH 7.4) and 0.1 M sorbitol for 1 h. Eyes were dehydrated and embedded in Spurr's resin. Ultrathin sections were post-stained with uranyl acetate and lead citrate. ZA were counted at the electron microscope. For the quantification of stalk length, we selected sections taken from 10 to 40 μ m below the cornea. Negatives taken at 3,000X were scanned and imported into Axiovision software (Zeiss) for measurements of stalk length. Statistics were calculated in Microsoft Excel. The same fixation was used for scanning electron microscopy. After dehydration, eyes were treated with hexamethyldisilazane, dried and sputter-coated with gold palladium.

Received 2 January; accepted 1 February 2002.

Published online 17 February 2002, DOI 10.1038/nature721.

- Nelson, W. J., Yeaman, C. & Grindstaff, K. K. in *Cell Polarity* (ed. Drubin, D. G.) 106–140 (Oxford Univ. Press, Oxford, 2000).
- Tepass, U., Tanentzapf, G., Ward, R. & Fehon, R. Epithelial cell polarity and cell junctions in *Drosophila*. *Annu. Rev. Genet.* **35**, 747–784 (2001).
- Dunn, R. E. in *The Ultrastructure of Sensory Organs* (ed. Friedmann, I.) 153–265 (Elsevier, New York, 1973).
- Wolff, T. & Ready, D. F. in *The Development of Drosophila melanogaster* (ed. Bate, M. Martinez Arias, A.) 1277–1325 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1993).
- Longley, R. L. Jr & Ready, D. F. Integrins and the development of three-dimensional structure in the *Drosophila* compound eye. *Dev. Biol.* **171**, 415–433 (1995).
- Tepass, U., Theres, C. & Knust, E. *crumbs* encodes an EGF-like protein expressed on apical membranes of *Drosophila* epithelial cells and required for organization of epithelia. *Cell* **61**, 787–799 (1990).
- Tepass, U. & Knust, E. Phenotypic and developmental analysis of mutations at the *crumbs* locus, a gene required for the development of epithelia in *Drosophila melanogaster*. *Roux's Arch. Dev. Biol.* **199**, 189–206 (1990).
- Wodarz, A., Hinz, U., Engelbert, M. & Knust, E. Expression of *Crumbs* confers apical character on plasma membrane domains of ectodermal epithelia of *Drosophila*. *Cell* **82**, 67–76 (1995).
- Tepass, U. *Crumbs*, a component of the apical membrane, is required for zonula adherens formation in primary epithelia of *Drosophila*. *Dev. Biol.* **177**, 217–225 (1996).
- den Hollander, A. I. et al. Mutations in a human homologue of *Drosophila crumbs* cause retinitis pigmentosa (RP12). *Nature Genet.* **23**, 217–221 (1999).

11. den Hollander, A. I. *et al.* Leber congenital amaurosis and retinitis pigmentosa with Coats-like exudative vasculopathy are associated with mutations in the *crumbs* homologue 1 (*CRB1*) gene. *Am. J. Hum. Genet.* **69**, 198–203 (2001).
12. Lotery, A. J. *et al.* Mutations in the *CRB1* gene cause Leber congenital amaurosis. *Arch. Ophthalmol.* **119**, 415–420 (2001).
13. den Hollander, A. I. *et al.* *CRB1* has a cytoplasmic domain that is functionally conserved between human and *Drosophila*. *Hum. Mol. Genet.* **10**, 2767–2773 (2001).
14. Bhat, M. A. *et al.* Discs Lost, a novel multi-PDZ domain protein, establishes and maintains epithelial polarity. *Cell* **96**, 633–645 (1999).
15. Klebes, A. & Knust, E. A conserved motif in *crumbs* is required for E-cadherin localization and zonula adherens formation in *Drosophila*. *Curr. Biol.* **10**, 76–85 (2000).
16. Bachmann, A., Schneider, M., Theilenberg, E., Grawe, F. & Knust, E. *Drosophila* Stardust is a partner of Crumbs in the control of epithelial cell polarity. *Nature* **414**, 638–643 (2001).
17. Hong, Y., Stronach, B., Perrimon, N., Jan, L. Y. & Jan, Y. N. *Drosophila* Stardust interacts with Crumbs to control polarity of epithelia but not neuroblasts. *Nature* **414**, 634–638 (2001).
18. Golc, K. G. Site-specific recombination between homologous chromosomes in *Drosophila*. *Science* **252**, 958–961 (1991).
19. Newsome, T. P., Asling, B. & Dickson, B. J. Analysis of *Drosophila* photoreceptor axon guidance in eye-specific mosaics. *Development* **127**, 851–860 (2000).
20. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
21. Tabuchi, K. *et al.* The GAL4/UAS-WGA system as a powerful tool for tracing *Drosophila* transsynaptic neural pathways. *J. Neurosci. Res.* **59**, 94–99 (2000).
22. Kumar, J. P. & Ready, D. F. Rhodopsin plays an essential structural role in *Drosophila* photoreceptor development. *Development* **121**, 4359–4370 (1995).
23. Grawe, F., Wodarz, A., Lee, B., Knust, E. & Skaer, H. The *Drosophila* genes *crumbs* and *stardust* are involved in the biogenesis of adherens junctions. *Development* **122**, 951–959 (1996).
24. Thomas, G. H. *et al.* *Drosophila* β_{Heavy} -spectrin is essential for development and contributes to specific cell fates in the eye. *Development* **125**, 2125–2134 (1998).
25. Thomas, G. Spectrin: The ghost in the machine. *Bioessays* **23**, 152–160 (2001).
26. Müller, H. A. & Wieschaus, E. *armadillo*, *bazooka*, and *stardust* are critical for early stages in formation of the zonula adherens and maintenance of the polarized blastoderm epithelium in *Drosophila*. *J. Cell Biol.* **134**, 149–163 (1996).
27. Wodarz, A., Ramrath, A., Grimm, A. & Knust, E. *Drosophila* atypical protein kinase C associates with Bazooka and controls polarity of epithelia and neuroblasts. *J. Cell Biol.* **150**, 1361–1374 (2000).
28. Petronczki, M. & Knoblich, J. A. DmPAR-6 directs epithelial polarity and asymmetric cell division of neuroblasts in *Drosophila*. *Nature Cell Biol.* **3**, 43–49 (2001).
29. Milam, A. H., Li, Z. Y. & Fariss, R. N. Histopathology of the human retina in retinitis pigmentosa. *Prog. Retinal Eye Res.* **17**, 175–205 (1998).
30. van Soest, S., Westerveld, A., de Jong, P. T., Bleeker-Wagemakers, E. M. & Bergen, A. A. Retinitis pigmentosa: defined from a molecular point of view. *Surv. Ophthalmol.* **43**, 321–334 (1999).
31. Bessant, D. A., Ali, R. R. & Bhattacharya, S. S. Molecular genetics and prospects for therapy of the inherited retinal dystrophies. *Curr. Opin. Genet. Dev.* **11**, 307–316 (2001).
32. Bennett, V. & Gilligan, D. M. The spectrin-based membrane skeleton and micron-scale organization of the plasma membrane. *Annu. Rev. Cell Biol.* **9**, 27–66 (1993).
33. De Matteis, M. A. & Morrow, J. S. Spectrin tethers and mesh in the biosynthetic pathway. *J. Cell Sci.* **113**, 2331–2343 (2000).
34. Morrow, J. S., *et al.* in *Handbook of Physiology* (eds Hoffman, J. & Jamieson, J.) 485–540 (Oxford Univ. Press, Oxford, 1997).
35. Kamal, A., Ying, Y. & Anderson, R. G. Annexin VI-mediated loss of spectrin during coated pit budding is coupled to delivery of LDL to lysosomes. *J. Cell Biol.* **142**, 937–947 (1998).
36. Michaely, P., Kamal, A., Anderson, R. G. & Bennett, V. A requirement for ankyrin binding to clathrin during coated pit budding. *J. Biol. Chem.* **274**, 35908–35913 (1999).
37. Rebay, I. *et al.* Specific EGF repeats of Notch mediate interactions with Delta and Serrate: implications for Notch as a multifunctional receptor. *Cell* **67**, 687–699 (1991).
38. Talts, J. F., Andac, Z., Gohring, W., Brancaccio, A. & Timpl, R. Binding of G domains of Laminin $\alpha 1$ and $\alpha 2$ chains and perlecan to heparin, sulfatides, α -dystroglycan and several extracellular matrix proteins. *EMBO J.* **18**, 863–870 (1999).
39. Bascom, R. A. Cloning of the cDNA for a novel photoreceptor membrane protein (rom-1) identifies a disk rim protein family implicated in human retinopathies. *Neuron* **8**, 1171–1184 (1992).
40. Arikawa, K., Molday, L. L., Molday, R. S. & Williams, D. S. Localization of peripherin/rds in the disk membranes of cone and rod photoreceptors: relationship to disk membrane morphogenesis and retinal degeneration. *J. Cell Biol.* **116**, 659–667 (1992).
41. Tanentzapf, G., Smith, C., McGlade, J. & Tepass, U. Apical, lateral, and basal polarization cues contribute to the development of the follicular epithelium during *Drosophila* oogenesis. *J. Cell Biol.* **151**, 891–904 (2000).
42. Byers, T. J., Dubreuil, R., Branton, D., Kiehart, D. P. & Goldstein, L. S. *Drosophila* spectrin. II. Conserved features of the α -subunit are revealed by analysis of cDNA clones and fusion proteins. *J. Cell Biol.* **105**, 2103–2110 (1987).
43. Thomas, G. H. & Kiehart, D. P. β_{Heavy} -spectrin has a restricted tissue and subcellular distribution during *Drosophila* embryogenesis. *Development* **120**, 2039–2050 (1994).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We are grateful to H. Hong and R. Derrane for technical assistance. We thank I. Dworkin for advice on statistical analysis. We thank H. Okano, the Developmental Studies Hybridoma Bank, and the Bloomington *Drosophila* Stock Center for reagents. We thank A. Satoh for permission to use the micrograph in Fig. 1g. We are grateful to S. Izaddoost and K.-W. Choi for discussing results before publication. We thank D. Godt, H. Lipshitz and E. Larsen for their comments on the manuscript. This work was supported by grants from the Canadian Institute of Health Research (to U.T.), the National Cancer Institute of Canada (to C.J.M.), and the National Institute of Health (to D.F.R.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to U.T. (e-mail: utepass@zoo.utoronto.ca).

A velocity dipole in the distribution of radio galaxies

Chris Blake & Jasper Wall

Astrophysics, Nuclear and Astrophysics Laboratory, University of Oxford, Keble Road, Oxford OX1 3RH, UK

The motion of our Galaxy through the Universe is reflected in a systematic shift in the temperature of the cosmic microwave background¹—because of the Doppler effect, the temperature of the background is about 0.1 per cent higher in the direction of motion, with a correspondingly lower temperature in the opposite direction. This effect is known as dipole anisotropy. If our standard cosmological model is correct, a related dipole effect should also be present as an enhancement in the surface density of distant galaxies in the direction of motion². The main obstacle to finding this signal is the uneven distribution of galaxies in the local supercluster, which drowns out the small cosmological signal. Here we report a detection of the expected cosmological dipole anisotropy in the distribution of galaxies. We use a survey of radio galaxies that are mainly located at cosmological distances, so the contamination from nearby clusters is small. When local radio galaxies are removed from the sample, the resulting dipole is in the same direction as the temperature anisotropy of the microwave background, and close to the expected amplitude. The result therefore confirms the standard cosmological interpretation of the microwave background.

The dipole anisotropy in the galaxy distribution due to our motion—the ‘velocity dipole’—has two causes. First, emissions

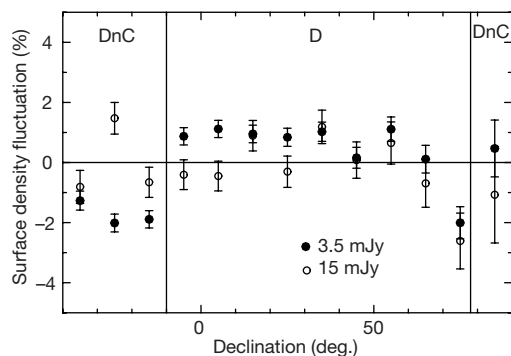


Figure 1 The NVSS suffers from systematic fluctuations in source surface density across the sky at low flux densities. Here we plot the surface density as a function of declination, in declination bins of width 10° , for flux-density thresholds 3.5 mJy (filled circles) and 15 mJy (open circles). The declination range of each array configuration (D or DnC) is shown. The 2% fluctuations present at a completeness limit of 3.5 mJy largely disappear when the threshold is raised to 15 mJy. These biases result from the sparse uv -plane sampling of the NVSS (W. Cotton, personal communication). The fluctuations are a function of declination only, as the NVSS beam and analysis procedure have no way of knowing the right ascension of objects. The NVSS was carried out with the Very Large Array at 1.4 GHz in D and DnC configurations during 1993–96. The survey covers all the sky north of declination -40° ; the catalogue contains $\sim 1.8 \times 10^6$ sources, and is claimed to be $\sim 99\%$ complete at integrated flux density 3.5 mJy. The full-width at half-maximum of the synthesized beam is ~ 45 arcsec. Before analysis, we masked 22 regions around bright and extended radio galaxies; these can appear in the NVSS catalogue as clusters of many point sources by virtue of the source-finding algorithm. These regions constitute a tiny fraction of the surveyed area and were filled with random distributions of sources at the mean surface density of the rest of the survey. The large NVSS beam means that only a small fraction of galaxies are resolved into double sources—this is important because a large number of doubles would increase the source surface density and hence the observed dipole amplitude.

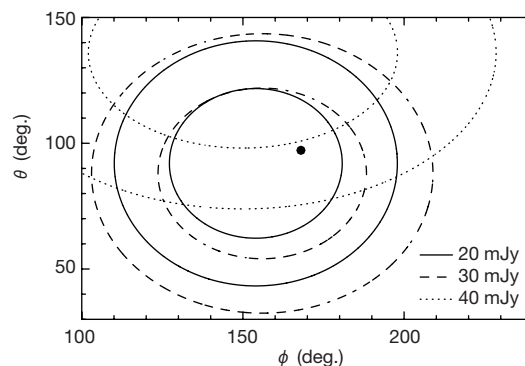


Figure 2 Our measured dipole direction agrees with that of the CMB dipole. Here we plot 1σ and 2σ contours for the best-fitting dipole direction for flux-density thresholds of 40 mJy (dotted lines), 30 mJy (dashed) and 20 mJy (solid). These 1σ and 2σ contours are defined by $\Delta\chi^2 = 2.30$ and 6.17 (with θ and ϕ varying, and δ fixed at its best-fitting value). The filled circle illustrates the direction of the CMB dipole, which is consistent with our measurements.

from galaxies ahead of us are Doppler boosted with respect to those behind, enhancing their fluxes; hence more are detected above a given flux threshold. Secondly, aberration of angles (as predicted by special relativity) causes the position vectors of galaxies to be displaced towards our direction of motion. We need to distinguish this velocity dipole—an imprint in the distribution of distant, evenly distributed galaxies—from the anisotropic distribution of local galaxies due to the local supercluster. We describe the latter as the ‘clustering dipole’; it will not be a dipole distribution, but will have a dipole component. The clustering dipole is the cause of our motion through the Universe (the gravitational pull of the local supercluster), while the velocity dipole is the consequence of it. To measure the velocity dipole we must remove the contamination of the clustering dipole.

The velocity dipole is a small effect, causing a surface density enhancement of $\sim 1\%$. In order to detect it, we must identify a sufficient quantity of distant objects and maintain uniformity in calibration to better than 1% over the sky. Active galactic nuclei (AGN) are situated in galaxies at cosmological distances, and are therefore ideal probes of the velocity dipole. Unsuccessful attempts

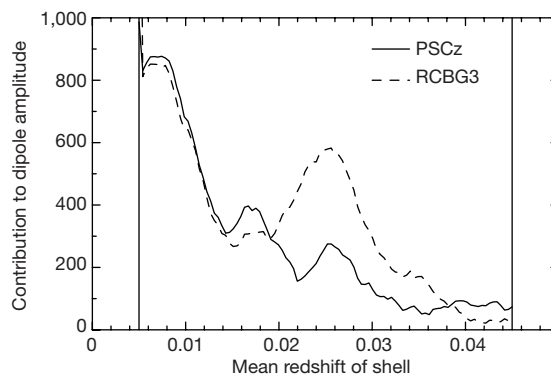


Figure 3 The clustering dipole originates from redshifts $z < 0.03$. Here we measure the strength of the clustering dipole in the PSCz and RCBG3 catalogues (see text) as a function of redshift. At each redshift z , we average over a redshift range $\Delta z = 0.01$ and thus the plot starts at $z = 0.005$. For each redshift shell, the y -axis plots $\sqrt{D^2 - N}$ where $D = |\sum \hat{r}_i|$ for the galaxies in that shell⁴, $\hat{r}_i$ being the unit position vector on the sky of the i th galaxy. This expression is proportional to the imbalance in the total number of sources across the sky, the quantity to which the total dipole amplitude is sensitive. The peak in RCBG3 at $z = 0.025$ is due to the Coma cluster. Hence it seems that once we get past Coma, the local clustering dipole is negligible.

have been made to detect this dipole in AGN X-ray surveys³ and radio surveys⁴. In contrast, the clustering dipole has been detected and quantified in studies probing the local distribution of galaxies at both optical⁵ and infrared⁶ wavelengths. In particular, the IRAS Infrared Astronomical Satellite PSCz catalogue has been used⁶ to compute the implied motion of the Local Group of galaxies.

We measure the galaxy dipole using the NRAO VLA Sky Survey of radio sources (NVSS)⁷. The NVSS covers a large fraction of the celestial sphere (82%) and is deep (~ 50 sources per degree²), thus providing sufficient sources for good statistics. Moreover, the radio AGN in the survey are at a median redshift $\bar{z} \approx 1$ (ref. 8), so that the contamination from local sources should be minimal. To prevent pollution from Galactic radio sources, we restrict our analysis to Galactic latitudes greater than 15° . The expected dipole amplitude is so small that systematic fluctuations above 1% over wide angles will drown the dipole signal. At flux densities below 15 mJy, the NVSS does suffer from spurious systematic fluctuations in source surface density of this magnitude (Fig. 1).

To eliminate the clustering dipole, we remove from the sample radio sources within 30 arcsec of known nearby galaxies as listed in the IRAS PSCz catalogue⁹ and the Third Reference Catalogue of Bright Galaxies (RCBG3)¹⁰. The IRAS galaxies eliminate local spiral galaxies obeying the far-infrared/radio correlation, and the RCBG3 sources remove radio sources powered by AGN in local elliptical galaxies.

To measure the dipole from the remaining distribution of galaxies, we measure the spherical harmonics of the distribution. Any density field across the sky $\sigma(\theta, \phi)$ —where $\theta(0 \rightarrow \pi)$ and $\phi(0 \rightarrow 2\pi)$ are ordinary spherical polar coordinates—can be expanded as a sum of spherical harmonics $Y_{lm}(\theta, \phi)$ with coefficients A_{lm} :

$$\sigma(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{+l} A_{lm} Y_{lm}(\theta, \phi) \quad (1)$$

The l th spherical harmonics produce fluctuations on angular scales $\sim 180^\circ/l$ and are described by $2l + 1$ independent coefficients. A dipole distribution across the sky will produce signal in the three $l = 1$ spherical harmonic coefficients, but no signal in the others. To estimate the coefficients A_{lm} from N discrete objects with positions (θ_i, ϕ_i) , we calculate:

$$A_{lm} = \sum_{i=1}^N Y_{lm}^*(\theta_i, \phi_i) \quad (2)$$

We measure the harmonic coefficients A_{lm} of the NVSS galaxy distribution for $1 \leq l \leq 3$ using equation (2) (a total of 15 independent coefficients). We measure the $l > 1$ coefficients because, as the measured sky is incomplete (containing blank regions below declination -40° and inside Galactic latitude 15°), higher harmonics are influenced by a dipole distribution. Further, any calibration gradients would (in general) bias all harmonics. Having measured the coefficients, we then determine how well they are fitted by a dipole model (with the same blank regions).

A dipole model has three parameters: an amplitude δ and direction described by two angles (θ, ϕ) . We define δ as the total fractional fluctuation in surface density from maximum to minimum. Table 1 displays the best-fitting dipole parameters and errors for a sequence of flux-density thresholds, and describes the method used. The dipole model is a good fit ($\chi^2_{\text{red}} \approx 1$) down to 15 mJy. Below this level the NVSS data-reduction biases of Fig. 1 become significant, producing a sudden increase in χ^2_{red} and a significant shift in the best-fitting θ (but not ϕ), anticipated because the biases are declination-dependent. Figure 2 illustrates the best-fitting direction (θ, ϕ) of the measured dipole at intensity levels above 20 mJy, which does not change with flux-density threshold, and is in good agreement with the direction of the cosmic microwave background (CMB) velocity dipole $\theta = 97.2^\circ \pm 0.1^\circ$, $\phi = 168.0^\circ \pm 0.1^\circ$.

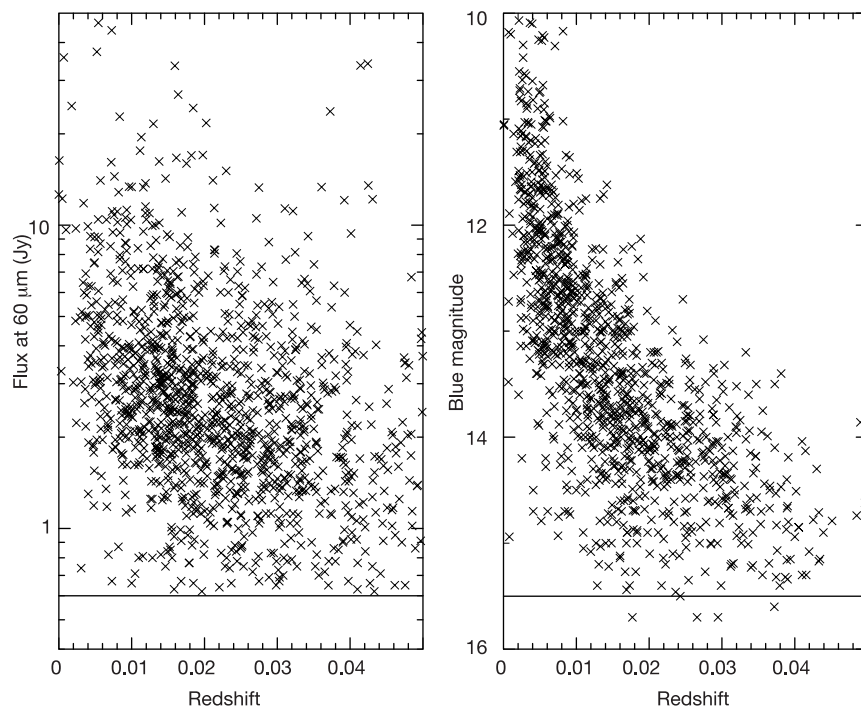


Figure 4 The PSCz and RCBG3 catalogues include almost all nearby NVSS galaxies. Here we present intensity–redshift plots for PSCz galaxies (left-hand panel, intensity in Jy) and for RCBG3 galaxies (right-hand panel, intensity in magnitudes) matched with NVSS sources brighter than 15 mJy. The horizontal lines are the stated completeness limits for the two catalogues. Clearly the bulk of the radio galaxies lie safely above the completeness

limits up to at least $z = 0.03$. Although the number of RCBG3 matches starts to die away at $z \approx 0.025$, this is probably due to the redshift distribution of the radio population rather than any serious incompleteness in RCBG3—doing the same plot for all RCBG3 galaxies shows no such fall-off until the magnitude threshold is reached.

Table 1 The dipole parameters best fitting the galaxy distribution

Flux (mJy)	<i>N</i>	Best δ ($\times 10^{-2}$)	Best ϕ (deg.)	Best θ (deg.)	χ^2_{red}
>40	125,603	1.4 ± 0.9	149 ± 49	135 ± 38	1.02
>35	143,524	1.7 ± 0.7	161 ± 44	117 ± 39	0.74
>30	166,694	2.2 ± 0.7	156 ± 32	88 ± 33	1.01
>25	197,998	2.2 ± 0.6	158 ± 30	94 ± 34	1.01
>20	242,710	2.1 ± 0.6	153 ± 27	93 ± 29	1.32
>15	311,037	1.5 ± 0.4	148 ± 29	59 ± 31	1.81
>10	431,990	1.0 ± 0.4	132 ± 29	25 ± 19	4.96

The table lists the best-fitting dipole model parameters (δ, θ, ϕ) together with 1σ errors for NVSS sources above various flux thresholds, after local sources have been removed. δ is expressed as a percentage. N is the number of sources left in the unmasked survey region and the reduced χ^2 values of the fit to the observed survey harmonics are also shown. The expected (CMB) velocity dipole parameters are $\delta = 0.9 \times 10^{-2}$, $\phi = 168^\circ$, $\theta = 97^\circ$. To find the best-fitting parameters, we vary the values of (δ, θ, ϕ) over a grid. For each set of parameters, we randomly generate a dipole galaxy distribution and mask it in the same way as the real data (below declination -40° and inside Galactic latitude 15°). We then measure the A_{lm} values of this test distribution, averaging over 100 realizations to increase accuracy. We hence obtain a mean and standard deviation for each A_{lm} in the set corresponding to the parameters (δ, θ, ϕ). We can then compute the χ^2 statistic with the observed survey harmonics. The best-fitting dipole parameters are found by minimizing χ^2 ; this minimum value (goodness-of-fit) indicates how well the dipole model fits the data. The errors on the best-fitting parameters are properly described by χ^2 contours in the parameter space. For example, with all three parameters varying, the 1σ and 2σ error regions are enclosed by χ^2 increases of 3.53 and 8.02.

(ref. 11). The local clustering dipole typically contributes $\Delta\delta = 0.5 \times 10^{-2}$ to the total amplitude.

The expected amplitude of the velocity dipole, assuming radio sources to have identical power-law spectra with flux density S depending on frequency ν as $S(\nu) \propto \nu^{-\alpha}$ and an integral source-count with N sources above S such that $N(>S) \propto S^{-x}$, is given by²:

$$\delta_{\text{pred}} = 2(u/c)[2 + x(1 + \alpha)] \quad (3)$$

where u is the peculiar velocity of our reference frame with respect to the frame in which the radio source population is (assumed) isotropic, and c is the speed of light. CMB dipole measurements imply $u = 370 \pm 2 \text{ km s}^{-1}$ (ref. 11). Taking $x \approx 1$ and a mean spectral index $\alpha \approx 0.75$, we find $\delta_{\text{pred}} = 0.9 \times 10^{-2}$. Note that u is the peculiar velocity of the Sun in the CMB frame. (The speed of the Earth around the Sun is $\sim 30 \text{ km s}^{-1}$ and can be neglected.) Our observed values of δ are on average 1.5σ away from this prediction.

We can be confident that our measured surface-density dipole is due to velocity rather than residual local clustering. Almost all the contribution to the local clustering dipole comes from redshifts $z < 0.03$ —analysis of the IRAS PSCz dipole⁶ shows that the dipole amplitude has almost converged by 100 Mpc. This result is verified in Fig. 3, where we plot the strength of the dipole contribution in redshift shells from the PSCz and RCBG3 catalogues. Furthermore, these catalogues contain almost all the radio galaxies to $z = 0.03$. This is demonstrated by Fig. 4, which gives magnitude–redshift plots for PSCz and RCBG3 galaxies matched with NVSS sources. The distributions make it clear that we are missing a negligible fraction of radio galaxies at $z < 0.03$.

It is hard to conceive of any other effect that could produce our measured dipole. A signal originating from instrumental or data-reduction effects would tend to skew the dipole direction towards the north or south poles, as such effects generally depend only on declination. However, our measured dipole points to declination approximately equal to 0° , in agreement with the CMB dipole. Moreover, instrumental effects should produce signal in higher (non-dipole) harmonics, but the best-fit dipole model is a good fit with $\chi^2_{\text{red}} \approx 1$. Furthermore, the dipole amplitude is extremely unlikely to be a chance fluctuation of an isotropic universe. For example, for a flux-density threshold of 20 mJy, the isotropic universe ($\delta = 0$) model fits the observed harmonics with $\chi^2_{\text{red}} \approx 2.28$, rejecting this model at a confidence level of $>99.5\%$. It is difficult to see what else could produce the dipole, particularly one pointing (within the 1σ direction contours) in the CMB dipole direction. For the 20 mJy measurement, the 1σ confidence region for the direction encompasses just 5% of the sky.

The alignment of our measured velocity dipole with the CMB dipole supports the predictions of fundamental assumptions about the cosmological origin of the CMB, the linchpin of much present-day observational cosmology. This result indicates that objects at $z \approx 1$ (such as distant radio galaxies) can define the cosmic frame, and have an isotropic distribution in that frame. $\square$

Received 11 October 2001; accepted 16 January 2002.

1. Smoot, G., Gorenstein, M. & Muller, R. Detection of anisotropy in the cosmic black-body radiation. *Phys. Rev. Lett.* **39**, 898–901 (1977).
2. Ellis, G. & Baldwin, J. On the expected anisotropy of radio source counts. *Mon. Not. R. Astron. Soc.* **206**, 377–381 (1984).
3. Scharf, C. *et al.* The 2–10 keV X-ray background dipole and its cosmological implications. *Astrophys. J.* **544**, 49–62 (2000).
4. Baleisis, A., Lahav, O., Loan, A. J. & Wall, J. V. Searching for large-scale structure in deep radio surveys. *Mon. Not. R. Astron. Soc.* **297**, 545–558 (1998).
5. Lahav, O. Optical dipole anisotropy. *Mon. Not. R. Astron. Soc.* **225**, 213–220 (1987).
6. Rowan-Robinson, M. *et al.* The IRAS PSCz dipole. *Mon. Not. R. Astron. Soc.* **314**, 375–397 (2000).
7. Condon, J. J. The NRAO VLA Sky Survey. *Astron. J.* **115**, 1693–1716 (1998).
8. Dunlop, J. S. & Peacock, J. A. The redshift cut-off in the luminosity function of radio galaxies and quasars. *Mon. Not. R. Astron. Soc.* **247**, 19–42 (1990).
9. Saunders, W. *et al.* The PSCz catalogue. *Mon. Not. R. Astron. Soc.* **317**, 55–64 (2000).
10. Corwin, H., Buta, R. & de Vacouleurs, G. Corrections and additions to the third reference catalogue of bright galaxies. *Astron. J.* **108**, 2128–2144 (1994).
11. Lineweaver, C. in *Proc. XVIth Moriond Astrophysics Meeting 69–75* (Editions Frontières, Gif-sur-Yvette, 1997).

Acknowledgements

We thank L. Miller and S. Rawlings for comments on early drafts. We also thank J. Condon for comments and suggestions.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.B. (e-mail: cab@astro.ox.ac.uk).

Electrical discharge from a thundercloud top to the lower ionosphere

Victor P. Pasko^{*†}, Mark A. Stanley^{†‡}, John D. Mathews^{*}, Umran S. Inan[§] & Troy G. Wood[§]

^{*} CSSL Laboratory, Penn State University, University Park, Pennsylvania 16802, USA

[†] Langmuir Laboratory, New Mexico Tech, Socorro, New Mexico 87801, USA

[§] STAR Laboratory, Stanford University, Stanford, California 94305, USA

[‡] These authors contributed equally to this work

For over a century, numerous undocumented reports have appeared about unusual large-scale luminous phenomena above thunderclouds^{1–6} and, more than 80 years ago, it was suggested that an electrical discharge could bridge the gap between a thundercloud and the upper atmosphere^{7,8}. Since then, two classes of vertically extensive optical flashes above thunderclouds have been identified—sprites^{9–11} and blue jets^{12–14}. Sprites initiate near the base of the ionosphere, develop very rapidly downwards at speeds which can exceed 10^7 m s^{-1} (ref. 15), and assume many

[‡] Present address: Space and Atmospheric Sciences, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

different geometrical forms^{16–19}. In contrast, blue jets develop upwards from cloud tops at speeds of the order of 10^5 m s^{-1} and are characterized by a blue conical shape^{12–14}. But no experimental data related to sprites or blue jets have been reported which conclusively indicate that they establish a direct path of electrical contact between a thundercloud and the lower ionosphere. Here we report a video recording of a blue jet propagating upwards from a thundercloud to an altitude of about 70 km, taken at the Arecibo Observatory, Puerto Rico. Above an altitude of 42 km—normally the upper limit for blue jets and the lower terminal altitude for sprites—the flash exhibited some features normally observed in sprites. As we observed this phenomenon above a relatively small thunderstorm cell, we speculate that it may be common and therefore represent an unaccounted for component of the global electric circuit.

The reported video recording was obtained during night-time observations using a Sony DCR TRV 730 charge-coupled device (CCD) video camera equipped with a blue extended ITT Night Vision GEN III NQ 6010 intensifier with a 40° circular field of view. The operation wavelength region of the intensifier was 390–870 nm at 77% sensitivity and 350–890 nm at 44% sensitivity. The intensifier provided a monochrome (predominantly green) image output. The camera was deployed at the Lidar Laboratory of Arecibo Observatory, Puerto Rico (18.347° N, 66.754° W, elevation 305 m above the sea level) (Fig. 1).

In the late evening of 14 September 2001, a cluster of thunderstorm cells developed at approximately 200 km range northwest of the observation site. Figure 1 shows an infrared image acquired from the GOES 8 satellite at 03:15 UT on 15 September 2001. The storm cloud top and the associated, almost continuous, lightning activity were clearly seen by the naked eye from Arecibo Observatory. The lower edge of the camera's field of view was aligned with the distant cloud top (altitude approximately 16 km) to avoid penetration of the direct light from bright lightning channels into the intensifier. The exact pointing direction of the video camera for the event was determined using the recorded star field (azimuth 336° , elevation 19.8° , corresponding to the centre of the camera field of view). GPS time stamps were recorded using the same video camera several hours before and several hours after the reported event, providing timing information with $\pm 33 \text{ ms}$ accuracy.

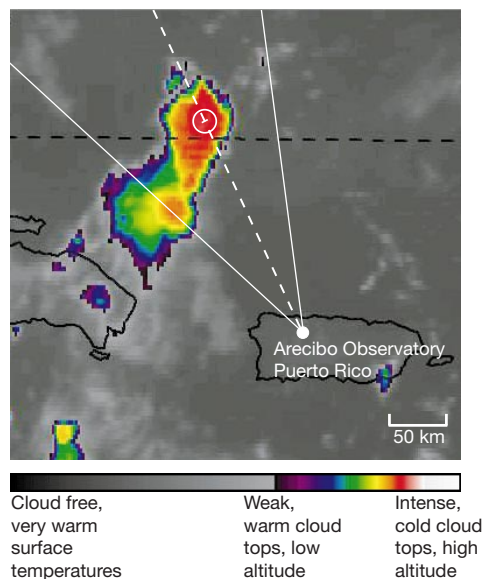


Figure 1 GOES-8 infrared image. This infrared image (1 km resolution) of the Puerto Rico region was acquired from the GOES-8 satellite at 03:15 UT on 15 September 2001 (<http://www.ghec.msfc.nasa.gov>).

The event was observed starting at 03:25:0.782 UT on 15 September 2001, lasted a total of 24 video frames ($\sim 33 \text{ ms}$ each), and concluded with an intense lightning flash in the underlying thundercloud in frame 25. Figure 2 shows a sequence of nine images extracted from frames 6–14. The full video sequence, including odd and even video fields for each frame, is available as Supplementary Information.

Figure 1 shows the camera's pointing direction and field of view. The dashed line represents the direction toward the apparent starting point of the event in frame 1 (see Supplementary Information), which intersects the coldest (and hence highest) cloud tops at $\sim 200 \text{ km}$ range. The 20-km-diameter circle drawn around the intersection point corresponds to the estimated transverse dimension (at ionospheric altitudes) of the observed phenomena. The maximum uncertainty in the estimated range of $\sim 200 \text{ km}$ is $\pm 20 \text{ km}$, corresponding to the width of the thunderstorm region, which translates to a $\pm 2 \text{ km}$ altitude uncertainty of features shown in Fig. 2, too small to affect any conclusions reported here.

The apparent speed of upward propagation of the observed phenomena remained remarkably stable during the first five frames, and is estimated to be $0.5 \times 10^5 \text{ m s}^{-1}$ ($\pm 0.07 \times 10^5 \text{ m s}^{-1}$), consistent with known speeds of the leader process in conventional lightning²⁰. The speed increased to $1.6 \times 10^5 \text{ m s}^{-1}$ between frames 5 and 6, and to $2.7 \times 10^5 \text{ m s}^{-1}$ between frames 6 and 7. The analysis of the two video fields corresponding to frame 8 indicates that the large altitude change between frames 7 and 8 happened in two steps. During the first field the left branch, clearly visible in frame 7, extended up to altitude $\sim 70 \text{ km}$, while during the second field the right branch formed with a wider tree-like structure. The altitude change of $\sim 32 \text{ km}$ for the left branch and $\sim 37 \text{ km}$ for the right branch happened faster than the duration of one video field (16.7 ms). The estimated speed in the range $(1.9\text{--}2.2) \times 10^6 \text{ m s}^{-1}$ is therefore a lower bound on the actual speed, which was probably higher.

The electromagnetic signatures of lightning discharges ('sferics') associated with the observed luminous event were recorded at two

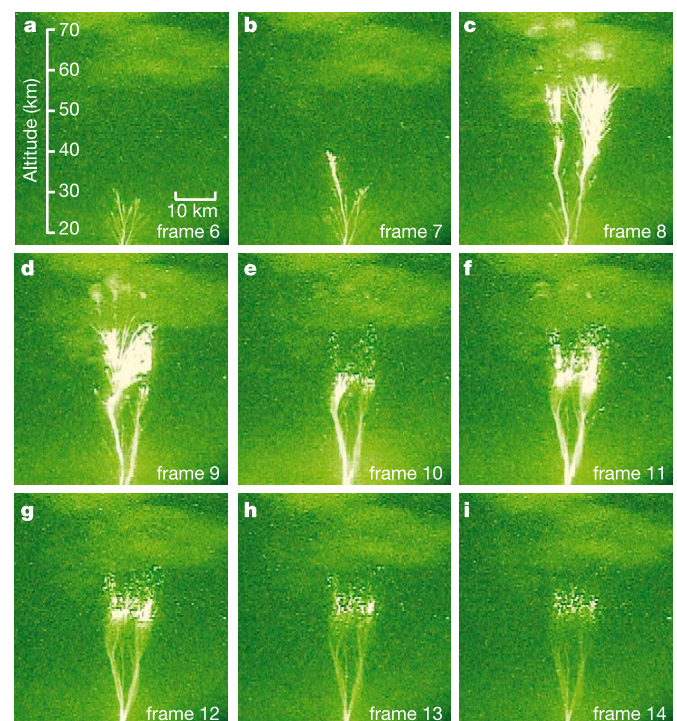


Figure 2 The time dynamics of the reported luminous event. Panels a–i correspond, respectively, to frames 6–14 as discussed in the text.

locations: Dominguito (18.421° N, 66.740° W), Puerto Rico, and at Palmer Station (64.774° S, 64.054° W), Antarctica. Dramatic rebrightening of the event was observed in frames 18 and 25 (see Supplementary Information) in association with large sferics. The sferic polarity indicated an upward transport of negative charge during the rebrightening. Also, an upward motion of subsequent breakdown was clearly evident between frames 18 and 19 (see Supplementary Information). As it is commonly known that subsequent breakdown of a given polarity in lightning generally occurs in the same direction as the initial breakdown of the same polarity, it is likely that the jet itself was created by upward negative breakdown, as was first suggested in ref. 28. A detailed report on the electromagnetic signatures will be presented elsewhere.

The apparent diameter of the breakdown filaments in the recorded images is estimated to be 1.26 km (± 0.26 km), similar to the estimated diameters of stars recorded in the same images. Thus our video recordings cannot be used to reliably estimate the minimum scale of the structure, although they provide conclusive evidence for the presence of filamentary branching structure as an essential component of the observed phenomena. This finding agrees with a recently reported colour photograph showing details of streamers in blue jets¹⁴, and provides the most detailed evidence presented to date of the theoretically predicted internal streamer structure of blue jets^{21–23}.

The transition apparent in frame 10 between the upper region dominated by hotspots and the lower region dominated by relatively smooth filamentary structure is estimated to be at an altitude of 42 km. This transition height is similar to the normal upper terminal altitude of blue jets and the lower terminal altitude of sprites.

The upward branches of the reported phenomena exhibit a diffuse termination at an altitude of approximately 70 km in frame 8. The 70 km termination closely coincides with the steep ledge of the lower ionospheric conductivity profile typically recorded during night-time rocket measurements in equatorial regions (for example, in Kenya²⁴ and Peru²⁵).

The initial stage of the observed phenomena closely resembles the general geometrical shapes and propagation speeds of previously documented blue jets^{12–14}, and the event was seen visually to be blue in colour. We therefore speculate that it can be classified as a blue jet, which propagated upward beyond the previously documented altitude. Our results represent, to our knowledge, the first video observation of blue-jet phenomena from the ground, although we emphasize that the possibility of such observations was demonstrated by previous ground-based video recordings of blue starters²⁶, which are probably related to the initial phases of blue jets^{23,27}.

The subsequent dynamics of the upper part of the phenomenon closely resembles some of the features often observed in sprites^{15–17} (that is, the shape of branching discharge trees, the diffuse termination of the breakdown branches on the lower ionospheric boundary, the evolution of the discharge trees into hotspots, and the high propagation speed). We note also that there are some similarities in visual appearance of the observed phenomenon and the so-called palm-tree events, which follow the occurrence of large groups of sprites and exhibit the same red colour as sprites^{18,19}. However, some other features of the observed phenomenon, such as its long duration, the altitude extent, and no apparent association with a positive cloud to ground lightning discharge, do not match typical properties of sprites.

The phenomenon we report here, which effectively demonstrates that an electrical contact between the thundercloud and the lower ionosphere was established in this event, may have been facilitated by the relatively low night-time middle atmospheric conductivity typically observed in the tropics^{24,25}. The lower conductivity values correspond to longer dielectric relaxation timescales (that is, ϵ_0/σ , where ϵ_0 is the permittivity of free space and σ is the conductivity), and therefore allow easier penetration of quasi-static thundercloud

electric fields—which presumably drive the observed phenomena²³—to higher altitudes. Given the fact that the phenomenon was produced by a relatively small ($\sim 2,500$ km²) thunderstorm cell, such cloud-to-ionosphere discharges may be very common in the tropics and may constitute an important, but as yet unaccounted for, component of the global electric circuit^{29,30}. □

Received 6 November 2001; accepted 24 January 2002.

- Wood, C. A. Unusual lightning. *Weather* **6**, 64 (1951).
- Wright, J. B. A thunderstorm in the tropics. *Weather* **6**, 230 (1951).
- Vaughan, O. H. Jr & Vonnegut, B. Lightning to the ionosphere? *Weatherwise* **70**–72 (April, 1982).
- Vaughan, O. H. Jr & Vonnegut, B. Recent observations of lightning discharges from the top of a thunderstorm into the clear air above. *J. Geophys. Res.* **94**, 13179–13182 (1989).
- Hammerstrom, J. G. Mystery lightning. *Aviat. Week Space Technol.* (30 August, 1993).
- Rodger, C. J. Red sprites, upward lightning, and VLF perturbations. *Rev. Geophys.* **37**, 317–336 (1999).
- Wilson, C. T. R. The electric field of a thundercloud and some of its effects. *Proc. Phys. Soc. Lond.* **37**, 32D–37D (1925).
- Wilson, C. T. R. A theory of thundercloud electricity. *Proc. R. Soc. Lond. A* **236**, 297–317 (1956).
- Franz, R. C., Nemzek, R. J. & Winckler, J. R. Television image of a large upward electrical discharge above a thunderstorm system. *Science* **249**, 48–51 (1990).
- Sentman, D. D., Wescott, E. M., Osborne, D. L., Hampton, D. L. & Heavner, M. J. Preliminary results from the Sprites94 campaign: Red Sprites. *Geophys. Res. Lett.* **22**, 1205–1209 (1995).
- Lyons, W. A. Sprite observations above the U.S. high plains in relation to their parent thunderstorm systems. *J. Geophys. Res.* **101**, 29641–29652 (1996).
- Wescott, E. M., Sentman, D. D., Osborne, D. L., Hampton, D. L. & Heavner, M. J. Preliminary results from the sprites 94 aircraft campaign: 2. Blue jets. *Geophys. Res. Lett.* **22**, 1209–1213 (1995).
- Wescott, E. M., Sentman, D. D., Heavner, M. J., Hampton, D. L. & Vaughan, O. H. Jr Blue jets: Their relationship to lightning and very large hailfall, and their physical mechanisms for their production. *J. Atmos. Solar-Terr. Phys.* **60**, 713–724 (1998).
- Wescott, E. M. *et al.* New evidence for the brightness and ionization of blue starters and blue jets. *J. Geophys. Res.* **106**, 21549–21554 (2001).
- Stanley, M. *et al.* High speed video of initial sprite development. *Geophys. Res. Lett.* **26**, 3201–3204 (1999).
- Gerken, E. A., Inan, U. S. & Barrington-Leigh, C. P. Telescopic imaging of sprites. *Geophys. Res. Lett.* **27**, 2637–2640 (2000).
- Stenbaek-Nielsen, H. C., Moudry, D. R., Wescott, E. M., Sentman, D. D. & Sao Sabbas, F. T. Sprites and possible mesospheric effects. *Geophys. Res. Lett.* **27**, 3827–3831 (2000).
- Heavner, M. J. *Optical Spectroscopic Observations of Sprites, Blue Jets, and Elves: Inferred Microphysical Processes and their Macrophysical Implications*. Thesis, Univ. Alaska (2000).
- Moudry, D. R., Stenbaek-Nielsen, H. C., Sentman, D. D. & Wescott, E. M. in *Abstr. National Radio Science Meeting of URSI G/H3-12*, 107 (Boulder, Colorado, 2001).
- Uman, M. A. *The Lightning Discharge* (Academic, San Diego, 1987).
- Petrov, N. I. & Petrova, G. N. Physical mechanisms for the development of lightning discharges between a thundercloud and the ionosphere. *Tech. Phys.* **44**, 472–475 (1999).
- Pasko, V. P., Inan, U. S. & Bell, T. F. Large scale modeling of sprites and blue jets. *Eos* **80**, F218 (1999).
- Pasko, V. P. & George, J. J. Three-dimensional modeling of blue jets and blue starters. *Eos* **82**, F150 (2001).
- Hale, L. C., Croskey, C. L. & Mitchell, J. D. Measurements of middle-atmosphere electric fields and associated electrical conductivities. *Geophys. Res. Lett.* **8**, 927–930 (1981).
- Croskey, C. L., Hale, L. C. & Mitchell, J. D. Electrical stratification in the middle atmosphere. *Adv. Space Res.* **10**, 49–52 (1990).
- Lyons, W. A., Stanley, M. A., Nelson, T. E. & Taylor, M. Sprites, elves, halos, trolls and blue starters above the STEPS Domain. *Eos* **81**, F131 (2000).
- Wescott, E. M. *et al.* Starters: Brief upward discharges from an intense Arkansas thunderstorm. *Geophys. Res. Lett.* **23**, 2153–2156 (1996).
- Sukhorukov, A. I., Mishin, E. V., Stubbe, P. & Rycroft, M. J. On blue jet dynamics. *Geophys. Res. Lett.* **23**, 1625–1628 (1996).
- Roble, R. G. On modeling component processes in the Earth's global electric circuit. *J. Atmos. Terr. Phys.* **53**, 831–847 (1991).
- Rycroft, M. J., Israelsson, S. & Price, C. The global atmospheric electric circuit, solar activity and climate change. *J. Atmos. Solar-Terr. Phys.* **62**, 1563–1576 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

The GEN III intensifier was provided by ITT Night Vision Industries; we thank M. Robinson for support of our work. We thank W. Lyons for discussions, and S. Gonzalez, Q. Zhou, M. Sulzer, C. Tepley, J. Friedman, E. Robles, A. Venkataraman and E. Castro for support of our observations at Arecibo Observatory. The Arecibo Observatory is a component of the National Astronomy and Ionosphere Center, which is operated by Cornell University under a cooperative agreement with the National Science Foundation. This work was supported by a Small Grant for Exploratory Research from the National Science Foundation to Pennsylvania State University. Stanford participation was also supported by the Office of Polar Programs of the National Science Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.P.P. (e-mail: vpasko@psu.edu) or M.A.S. (e-mail: stanley@m.lanl.gov).

Formation of isomorphous Ir³⁺ and Ir⁴⁺ octamers and spin dimerization in the spinel CuIr₂S₄

Paolo G. Radaelli^{*}, Y. Horibe^{†‡§}, Matthias J. Gutmann^{*}, Hiroki Ishibashi^{§||}, C. H. Chen[†], Richard M. Ibberson^{*}, Y. Koyama[‡], Yew-San Hor[§], Valery Kiryukhin[§] & Sang-Wook Cheong[§]

^{*} ISIS Facility, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, OX11 0QX, UK

[†] Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA

[‡] Department of Materials Science and Engineering and Kagami Memorial Laboratory for Materials Science and Technology, Waseda University, Shinjuku-ku, Tokyo 169, Japan

[§] Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854, USA

^{||} Department of Materials Science, Osaka Prefecture University, Sakai 599, Osaka, Japan

Inorganic compounds with the AB₂X₄ spinel structure have been studied for many years, because of their unusual physical properties. The spinel crystallographic structure, first solved by Bragg in 1915¹, has cations occupying both tetrahedral (A) and octahedral (B) sites. Interesting physics arises when the B-site cations become mixed in valence. Magnetite (Fe₃O₄) is a classic and still unresolved example, where the tendency to form ordered arrays of Fe²⁺ and Fe³⁺ ions competes with the topological frustration of the B-site network². The CuIr₂S₄ thiospinel is another example, well known for the presence of a metal–insulator transition at 230 K with an abrupt decrease of the electrical conductivity on cooling accompanied by the loss of localized magnetic moments^{3–7}. Here, we report the determination of the crystallographic structure of CuIr₂S₄ below the metal–insulator transition. Our results indicate that CuIr₂S₄ undergoes a simultaneous charge-ordering and spin-dimerization transition—a rare phenomenon in three-dimensional compounds. Remarkably, the charge-ordering pattern consists of isomorphous octamers of Ir₈³⁺S₂₄ and Ir₈⁴⁺S₂₄ (as iso-valent bi-capped hexagonal rings). This extraordinary arrangement leads to an elegant description of the spinel structure, but represents an increase in complexity with respect to all the known charge-ordered structures, which are typically based on stripes, slabs or chequerboard patterns.

Spin dimerization occurs when two neighbouring magnetic ions form a quantum superposition of states with the spins pointing in opposite directions (known as a spin singlet), so that the average magnetic moment is zero. In a periodic system, spin dimerization is always associated with a lattice distortion, leading to a stronger interaction of the spins within the dimers. However, spin–lattice dimerization is not expected to occur in a three-dimensional (3D) system such as CuIr₂S₄. In one-dimensional (1D) systems, a variety of mechanisms are known to promote spontaneous spin dimerization, the best known being coupling with soft phonons in the so-called spin-Peierls transitions^{8–10}. None of these mechanisms is operative for 3D systems, which, in the absence of irreconcilable interactions (frustration), are expected to order antiferromagnetically. Nevertheless, spin dimers could still form in three dimensions, should the lattice energy term favour short pair-wise distances or confine the spins to 1D topological structures within the lattice. To see if this dimerization has occurred, an accurate solution of the crystal structure is required. Because phase transitions in spinels can be very complex, we have tackled the problem with a variety of diffraction techniques.

Polycrystalline samples of CuIr₂S₄ were prepared by standard solid-state reaction. Electron diffraction data, high-resolution synchrotron

X-ray powder diffraction data and medium- and high-resolution neutron diffraction data were collected at various temperatures above and below the metal–insulator (MI) transition (see Methods). It has been previously shown³ that the MI transition is associated with a sharp first-order structural transition, lowering the crystal symmetry from the cubic (space group *Fd3m*) high-temperature phase. Furubayashi *et al.*³ proposed the tetragonal space group *I4₁/amd* for the low-temperature structure, but also noted the presence of unindexed Bragg peaks.

The electron diffraction patterns observed at room temperature from several different zone axes confirmed the cubic *Fd3m* space group. Below the MI transition (~230 K), new superlattice and forbidden spots appeared in addition to those allowed in the *I4₁/amd* space group (Fig. 1a). All these satellites, which are also present in the X-ray data collected at 50 K (Fig. 1b), can be indexed by two propagation vectors: [1/2,0,1/2] and [1,0,0] (*I4₁/amd* indexing), that is, the *N*- and *M*-points of the Brillouin zone, respectively. The overall lattice is triclinic¹¹, with lattice parameters *a* = 11.95032(10) Å, *b* = 6.98016(8) Å, *c* = 11.92776(9) Å, *α* = 91.0550(7)°, *β* = 108.4672(6)° and *γ* = 91.0320(7)° (see the inset in Fig. 1a).

By refining the medium-resolution neutron data in the *I4₁/amd* space group, we could estimate the size and direction of the atomic displacements. In the cubic phase, all the atomic displacements are small, but below 230 K, they become very large and anisotropic for iridium (the mean square shift is *U* ≈ 0.03 Å² along the Ir–Ir bonds), and, to a lesser extent, for sulphur. A starting model with 28 atoms in general positions of the asymmetric unit cell was set up for structural refinements within the triclinic space group *P1*. Displacement patterns for Ir and S were determined using the space-group representation theory (coupled *N*- and *M*-point representation). The directions and magnitudes of the eigenvectors were set consistently with the refined atomic displacements. A full, unconstrained refinement of the triclinic structure was performed using this starting model on the combined neutron and X-ray diffraction data. The overall quality of the fit is excellent: goodness-of-fit *χ*² = 3.0 and profile-weighted *R*-factor *R*_{wp} = 5.9% for the combined data. Portions of the Rietveld refinement plots for the synchrotron X-ray and medium-resolution neutron data are shown in Fig. 1b. The refined structure of CuIr₂S₄ is shown in Fig. 2a. The Ir–Ir bond lengths are reported in Table 1. In the Supplementary Information, we report the atomic parameters (Supplementary Table I), additional bond lengths (Supplementary Table II) and

Table 1 Ir–Ir bond lengths of CuIr₂S₄ at 50 K

Dimerized rings (Ir⁴⁺)

1:3	3.464(7)
1:4	3.012(5)
2:3	2.956(5)
2:4	3.582(7)
3:4	3.490(8)

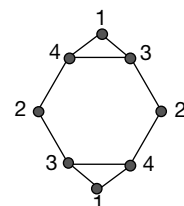
Non-dimerized rings (Ir³⁺)

1:3	3.478(7)
1:4	3.493(5)
2:3	3.602(5)
2:4	3.487(7)
3:4	3.470(9)

Ir⁴⁺–Ir³⁺ bonds across rings (Ir⁴⁺:Ir³⁺)

1:1	3.551(7)	2:4	3.605(7)
1:1	3.437(7)	3:1	3.658(5)
1:3	3.540(7)	3:2	3.458(8)
1:4	3.554(7)	3:3	3.458(7)
2:2	3.507(6)	4:1	3.486(7)
2:2	3.483(6)	4:2	3.581(5)
2:3	3.432(7)	4:4	3.554(8)

Data determined from Rietveld refinements of X-ray and neutron powder diffraction data. All bond lengths are given in Å.



the refined high-resolution neutron data (Supplementary Fig. 1).

We have employed the pair distribution function technique to obtain an independent confirmation of the correctness of the crystal structure. This procedure, which has been successfully employed in the past to determine local structures of both crystalline¹² and non-crystalline materials, is completely model-independent, and, unlike the Rietveld method, makes use of both Bragg peaks and diffuse scattering. In Fig. 3, we compare the pair distribution functions obtained from the measured data with the one calculated from the Rietveld-refined structural parameters. The agreement is remarkably good, a further indication of the correctness of the model in general and of the space group in particular.

A close examination of the structure in Fig. 2a reveals a remarkable complexity. There are eight independent Ir atoms in the

triclinic structure. Of these, four form very short metal–metal bonds (~ 3.0 Å) with each other across shared octahedral faces, all the other Ir–Ir distances being between 3.43 and 3.66 Å. It is impossible to determine the Ir valence reliably from the Ir–S bond lengths for such a complex structure where sulphur has a low scattering length for both X-rays and neutrons. Nonetheless, the mixed-valence nature of the Ir ions (Ir^{3+} and Ir^{4+}) has been established in a previous study^{4,13}, and it is natural to assume that the dimerized Ir atoms have a valence of 4+, the unpaired electrons in the d_{xy} orbitals forming singlet spin states, thus explaining the disappearance of the local magnetic moments. The two types of Ir—dimerized (Ir^{4+}) and non-dimerized (Ir^{3+})—are depicted in blue and red, respectively, in Fig. 2a. Atoms of each species are arranged in octamers—groups of 8 octahedra, related in pairs by

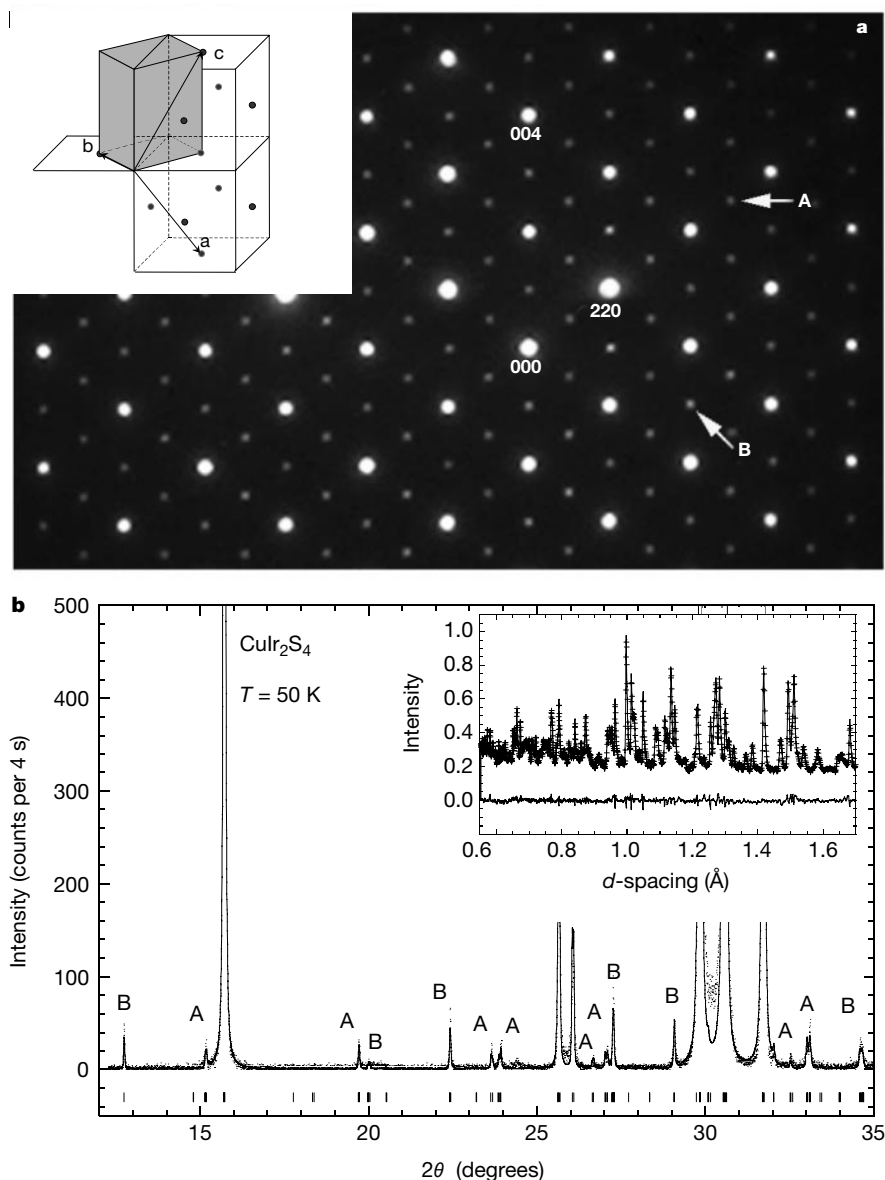


Figure 1 Superlattice peaks resulting from charge ordering and lattice dimerization in CuIr_2S_4 : electron, X-ray and neutron diffraction data. **a**, Main panel: electron diffraction pattern for CuIr_2S_4 at 93 K, with the incident beam parallel to the $[1, \bar{1}, 0]$ cubic zone axis. The spots labelled A double the tetragonal cell in two directions, and correspond to a $[1/2, 0, 1/2]$ -propagation vector (I_4/amd indexing), that is, the N -point of the Brillouin zone. Those labelled B violate the body-centring operator, and correspond to a $[1, 0, 0]$ -propagation vector, the M -point in the Brillouin zone. Inset: orientation of the triclinic cell

of CuIr_2S_4 at 50 K (axes are labelled and indicated with arrows) with respect to the room-temperature cubic ($Fd\bar{3}m$) spinel lattice. The dots indicate the face-centred Bravais lattice points. The tetragonal (I_4/amd) cell is shown in light grey. **b**, Main panel: portion of the Rietveld fit of the synchrotron X-ray powder diffraction pattern ($\lambda = 1.547$ Å) for CuIr_2S_4 at 50 K. Superstructure peaks are labelled as in **a**. Inset: portion of the Rietveld fit of the time-of-flight, medium-resolution neutron powder diffraction data for CuIr_2S_4 at 50 K.

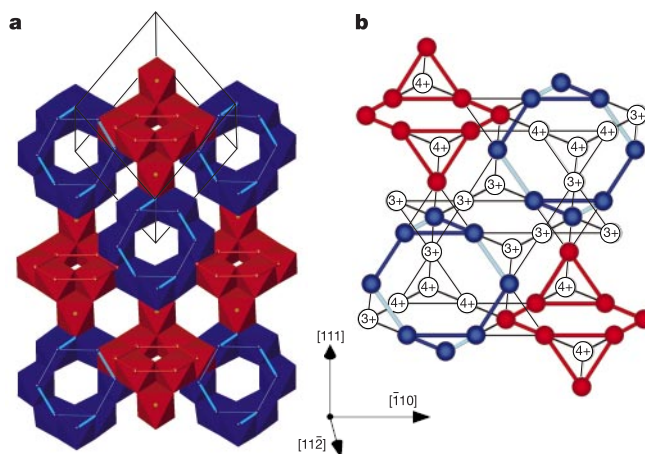


Figure 2 The low-temperature crystal structure of CuIr_2S_4 . **a**, Schematic representation of the crystal structure of CuIr_2S_4 at 50 K. Red and blue octahedra have Ir^{3+} and Ir^{4+} , respectively, at the centres, and S at the corners. For the latter, the dimerized Ir–Ir bonds are indicated with light-blue cylinders. The copper atoms (not shown) are in the tetrahedral holes of the octahedral framework. **b**, A view of the 3D configuration of the Ir ions, emphasizing the orientation of the hexagonal rings perpendicular to the $[111]$ cubic direction. The colour pattern and the projection are the same as for **a**. In the projection we employed, the $[111]$ cubic direction is tilted by approximately 18° out of the paper plane to

create a 3D impression of the stacking and to provide a prospective view of both types of octamers. The charge-ordering pattern is also indicated. The Ir^{3+} hexagons lie essentially in the (111) planes whereas the Ir^{4+} hexagons lie in the $(1,1,\bar{1})$ planes. Spin dimerization of the Ir^{4+} ions takes place along the two orthogonal cubic directions $[101]$ and $[1,0,\bar{1}]$. The other twin-related domains with hexagons in the $(1,\bar{1},1)/(\bar{1},1,1)$ planes have also been observed in our electron microscopy study. The electron diffraction pattern shown in Fig. 1a was obtained from a single twin domain.

the centre of symmetry. These can be described as planar hexagonal rings with two additional octahedra attached at opposite sides, above and below the hexagon (bi-capped hexagonal rings). The Ir ions (with the spin values of $1/2$) in the Ir^{4+} octamers exhibit drastic alternations of Ir–Ir distances ($\sim 3.0 \text{ \AA}$ and $\sim 3.5 \text{ \AA}$), resulting in spin dimerization, whereas the Ir–Ir distance in Ir^{3+} -octamers ($S = 0$) are uniform. This is further illustrated in Fig. 2b, which shows the undistorted 3D configuration of the Ir ions in the (111)

planes (cubic notation) for four octamers. It should be noted that the Ir^{3+} octamers and Ir^{4+} octamers are topologically similar, except for the internal dimerization pattern.

There exists a number of 1D $S = 1/2$ systems displaying spin-Peierls transitions coupled with lattice dimerization. Most of these compounds are organic¹⁰, but the first inorganic spin-Peierls system, CuGeO_3 , was discovered recently¹⁴. At low temperatures, Cu atoms in CuGeO_3 are dimerized¹⁵, with a distance modulation of

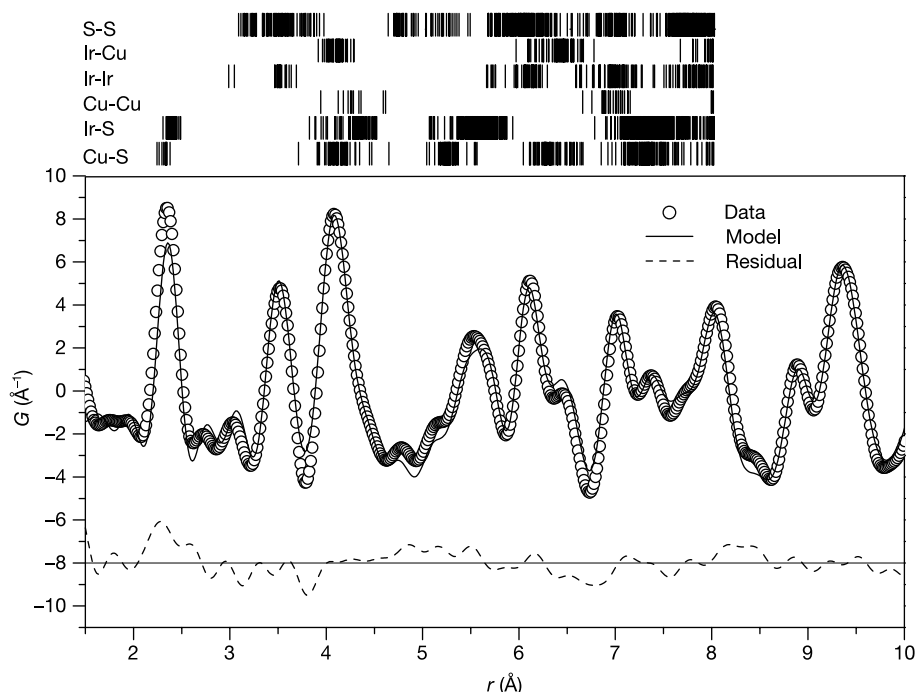


Figure 3 Pair distribution function of CuIr_2S_4 at 50 K. The model is calculated from the Rietveld-refined structure. Vertical lines above the data indicate contributions of the different atom pairs, with the following relative weights: Ir–Ir: 1.0; Ir–Cu: 0.728; Cu–Cu: 0.530; Ir–S: 0.269; Cu–S: 0.196; S–S: 0.072. Only a global atomic displacement

parameter for each of the species was refined, as these parameters are expected to be significantly different from the Rietveld ones, owing to the presence of correlated motion and/or local disorder. The refined values of the atomic displacement parameters are $13(2) \text{ pm}^2$, $17(2) \text{ pm}^2$ and $13(3) \text{ pm}^2$ for Ir, Cu and S, respectively.

approximately $\pm 6 \times 10^{-3}$ Å. The mixed-valence, quasi-two-dimensional compound α' - NaV_2O_5 is also thought to display a spin-Peierls transition at $T_{\text{sp}} = 33$ K, and is also charge ordered at low temperatures. However, α' - NaV_2O_5 is an insulator at room temperature, and its complex low-temperature displacement pattern¹⁶ cannot be easily interpreted in terms of dimerization. Ti_4O_7 , also a quasi-two-dimensional compound, is thought to undergo a spin-dimerization and charge-localization transition¹⁷. Other Magneli phases based on Ti and V may also display similar behaviour. Our results indicate that CuIr_2S_4 is a rare and possibly unique example of a 3D material displaying a spin-dimerization transition, which occurs simultaneously with charge ordering. This produces very large variations in the metal–metal bond lengths (± 0.25 Å), which are 40 times larger than for CuGeO_3 and about twice as large as in Ti_4O_7 .

We find that the peculiar form of charge ordering in CuIr_2S_4 is the most remarkable aspect of this material. The B-site (Ir ions) network in a spinel can be regarded as a diamond lattice of corner-sharing tetrahedra. In the past, most of the relevant physics was focused on these tetrahedra. In magnetite (Fe_3O_4), for example, it has been generally accepted that every tetrahedron satisfies Anderson's condition¹⁸, with two Fe^{2+} and two Fe^{3+} ions on the corners. The charge-ordering pattern in CuIr_2S_4 with octamers of Ir_8S_{24} octahedra offers a new way to look at the spinel structure, resulting in an elegant description based on a single topological building block. Here, not all the corner-sharing tetrahedra satisfy Anderson's condition for charge balance as can be seen in Fig. 2b. Given the unusually large Ir–Ir lattice distortions, it is not surprising that Coulomb interaction may not play the most significant role. Charge ordering in transition metal compounds invariably occurs in very simple geometrical patterns, based on stripes, columns, slabs or checkerboards—columnar charge ordering in spinel LiMn_2O_4 is a recent relevant example¹⁹. The topology of charge ordering in CuIr_2S_4 has, to our knowledge, no precedent in inorganic materials, and represents a step up in complexity with respect to known arrangements. The driving force leading to this complex pattern is yet to be understood; we hypothesize that the octamer configuration enables dimerized and non-dimerized pairs to coexist without perturbing each other in the highly interconnected spinel network. The prospect of determining the excitation spectrum of a $S = 1/2$ bi-capped hexagonal ring is of interest from both theoretical and experimental points of view. □

Methods

Sample preparation

A mixture of Cu, Ir and S powders was sealed in an evacuated quartz tube and was calcined at 850 °C for a period of 8 days. The resulting powder specimens were subsequently ground and pressed into pellets for a second sintering in a sealed quartz tube at 850 °C for 8 days. The sample was found to be single phase from X-ray diffraction analysis.

Sample characterization and electron diffraction

Resistivity was measured with the standard four-probe technique, and susceptibility measurements were performed using a Quantum Design SQUID magnetometer. These measurements were found to agree with literature data^{2–7}.

Both crushed and Ar-ion milled samples were used for the electron diffraction and transmission electron microscopy studies. No difference was observed between the two types of specimen. Our studies were carried out in a JEOL-2000FX transmission electron microscope, equipped with a low-temperature stage and a 14-bit charge-coupled device array of detectors.

X-ray and neutron powder diffraction

Synchrotron X-ray powder diffraction measurements were performed at the BL-3A beamline at the Photon Factory in Tsukuba, Japan. A complete data set was collected at 50 K, using a wavelength of 1.547 Å, and a series of scans of individual Bragg peaks were performed as a function of temperature. The beamline was equipped with a double Si(111) monochromator with horizontal (sagittal) focusing, and a Si(111) flat analyser, with the sample in flat-plate geometry. Under these conditions the 2θ resolution is 0.004° (full-width at half-maximum).

Medium- and high-resolution time-of-flight neutron powder diffraction data were

collected using the GEM and HRPD instruments, respectively, at the ISIS facility of the Rutherford Appleton Laboratory, UK. The GEM measurements were performed with the sample in an 8-mm-diameter cylindrical vanadium container, cooled by means of a closed-cycle refrigerator. Data were collected between 10 and 300 K at 10 K intervals, using a wavelength range of $0.3 \text{ Å} \leq \lambda \leq 4.5 \text{ Å}$, and focused in 4 groups at nominal scattering angles 2θ of 18°, 64°, 92° and 153°. The HRPD data were collected at a single temperature (50 K), using a wavelength range of $1.3 \text{ Å} \leq \lambda \leq 5 \text{ Å}$ with the sample in a slab geometry in a closed cycle refrigerator. We only employed the back-scattering data (nominal $2\theta = 168^\circ$) for the analysis.

Data analysis

Irreducible representation analysis of the atomic displacement modes was performed with the help of the program ISOTROPY²⁰. Rietveld refinements were performed simultaneously on the X-ray and neutron powder diffraction data at 50 K, for a total of 6 histograms, using the program GSAS²¹. The neutron-weighted experimental pair distribution function was calculated based on the merged GEM histograms using the program PDFgetN²². The program PDFFIT²³ was used to calculate the pair distribution function based on the structural parameters obtained from Rietveld refinements. No further refinement of the atomic coordinates was performed, but we refined an overall atomic displacement parameter for each of the species (see text).

Received 16 November 2001; accepted 29 January 2002.

1. Bragg, W. H. The structure of magnetite and the spinels. *Nature* **95**, 561–561 (1915).
2. Iizumi, M. *et al.* Structure of magnetite (Fe_3O_4) below the Verwey transition temperature. *Acta Crystallogr.* **38**, 2121–2133 (1982).
3. Furubayashi, T., Matsumoto, T., Hagino, T. & Nagata, S. Structural and magnetic studies of metal-insulator transition in thiospinel CuIr_2S_4 . *J. Phys. Soc. Jpn* **63**, 3333–3339 (1994).
4. Matsuno, J. *et al.* Photoemission study of the metal-insulator transition in CuIr_2S_4 . *Phys. Rev. B* **55**, R15979–R15982 (1997).
5. Nagata, S. *et al.* Metal-insulator transition in the spinel-type $\text{CuIr}_2(\text{S}_{1-x}\text{Se}_x)_4$ system. *Phys. Rev. B* **58**, 6844–6854 (1998).
6. Matsumoto, N., Endoh, R., Nagata, S., Furubayashi, T. & Matsumoto, T. Metal-insulator transition and superconductivity in the spinel-type $\text{Cu}(\text{Ir}_{1-x}\text{Rh}_x)_2\text{S}_4$ system. *Phys. Rev. B* **60**, 5258–5265 (1999).
7. Burkov, A. T. *et al.* Anomalous resistivity and thermopower of the spinel-type compounds CuIr_2S_4 and CuIr_2Se_4 . *Phys. Rev. B* **61**, 10049–10056 (2000).
8. Bray, J. W. *et al.* Observation of a spin-Peierls transition in a Heisenberg antiferromagnetic linear-chain system. *Phys. Rev. Lett.* **35**, 744–747 (1975).
9. Jacobs, I. S. *et al.* Spin-Peierls transitions in magnetic donor-acceptor compounds of tetrathiafulvalene (TTF) with bisdithiolenes metal complexes. *Phys. Rev. B* **14**, 3036–3051 (1976).
10. Bray, J. W., Interrante, I. V., Jacobs, I. S. & Bonner, J. C. *Extended Linear Chain Compounds* (ed. Miller, J. S.) (Plenum, New York, 1983).
11. Ishibashi, H., Sakai, T. & Nakahigashi, K. X-ray diffraction study on spinel compound CuIr_2S_4 with metal-insulator transition. *J. Magn. Magn. Mater.* **226–230**, 233–234 (2001).
12. Billinge, S. J. L. & Egami, T. Short-range atomic structure of $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_{4-x}$ determined by real-space refinement of neutron-powder-diffraction data. *Phys. Rev. B* **47**, 14386–14406 (1993).
13. Kumagai, K., Kakuyanagi, K., Endoh, R. & Nagata, S. NMR study on the metal-insulator transition of $(\text{Cu}_{1-x}\text{Ni}_x)\text{Ir}_2\text{S}_4$. *Physica C* **341**, 741–742 (2000).
14. Hase, M., Terasaki, I. & Uchinokura, K. Observation of the spin-Peierls transition in linear Cu^{2+} (spin-1/2) chains in an inorganic compound CuGeO_3 . *Phys. Rev. Lett.* **70**, 3651–3654 (1993).
15. Hirota, K. *et al.* Dimerization of CuGeO_3 in the Spin-Peierls state. *Phys. Rev. Lett.* **73**, 736–739 (1994).
16. Lüdecke, J. *et al.* Acentric low-temperature superstructure of NaV_2O_5 . *Phys. Rev. Lett.* **82**, 3633–3636 (1999).
17. Marezio, M., McWhan, D. B., Derniew, P. D. & Remeika, J. P. Charge localization at metal-insulator transitions in Ti_4O_7 . *Phys. Rev. Lett.* **28**, 1390–1393 (1972).
18. Anderson, P. W. Ordering and antiferromagnetism in ferrites. *Phys. Rev.* **102**, 1008–1013 (1956).
19. Rodríguez-Carvajal, J., Rouse, G., Masquelier, C. & Hervieu, M. Electronic crystallization in a lithium battery material: columnar ordering of electrons and holes in the spinel LiMn_2O_4 . *Phys. Rev. Lett.* **81**, 4660–4663 (1998).
20. Stokes, H. T. & Hatch, D. M. *Isotropy Subgroups of the 230 Crystallographic Space Groups* (World Scientific, Singapore, 1988).
21. Larson, A. C. & von Dreele, R. B. (The Regents of the University of California, Los Alamos, 1985).
22. Peterson, P. F., Gutmann, M., Proffen, T. & Billinge, S. J. L. PDFgetN: a user-friendly program to extract the total scattering structure factor and the pair distribution function from neutron powder diffraction data. *J. Appl. Crystallogr.* **33**, 1192 (2000).
23. Proffen, T. & Billinge, S. J. L. PDFFIT, a program for full profile structural refinement of the atomic pair distribution function. *J. Appl. Crystallogr.* **32**, 572–575 (1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

Y.-S.H., V.K. and S.-W.C. were partially supported by the National Science Foundation; Y.H. acknowledges the support of a JSPS Postdoctoral Fellowship for Research Abroad.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.G.R. (e-mail: p.g.radaelli@rl.ac.uk).

Onset of Asian desertification by 22 Myr ago inferred from loess deposits in China

Z. T. Guo*, William F. Ruddiman†, Q. Z. Hao*, H. B. Wu*, Y. S. Qiao*, R. X. Zhu*, S. Z. Peng*, J. J. Wei*, B. Y. Yuan* & T. S. Liu*

* Institute of Geology and Geophysics, Chinese Academy of Sciences, PO Box 9825, Beijing 100029, China

† Department of Environmental Science, Clark Hall, University of Virginia, Charlottesville, Virginia 22903, USA

The initial desertification in the Asian interior is thought to be one of the most prominent climate changes in the Northern Hemisphere during the Cenozoic era^{1–4}. But the dating of this transition is uncertain, partly because desert sediments are usually scattered, discontinuous and difficult to date. Here we report nearly continuous aeolian deposits covering the interval from 22 to 6.2 million years ago, on the basis of palaeomagnetic measurements and fossil evidence. A total of 231 visually definable aeolian layers occur as brownish loesses interbedded with reddish soils. This new evidence indicates that large source areas of aeolian dust and energetic winter monsoon winds to transport the material must have existed in the interior of Asia by the early Miocene epoch, at least 14 million years earlier than previously thought^{3,5}. Regional tectonic changes and ongoing global cooling are probable causes of these changes in aridity and circulation in Asia.

Uplift of the Himalayan–Tibetan plateau^{1–3} and changes in land–sea distribution⁴ have been invoked as driving forces behind long-term Cenozoic climate deterioration. One problem in understanding these links is defining the timing of major tectonic and climatic changes, including the onset and development of aridification in Asia. Continuous long-term records with well constrained chronologies are especially important.

Desert loess is particularly valuable as an indicator of dryland evolution because sizeable deserts are needed as sources for aeolian deposits^{5–7}. As a result, most studies focus on regions which receive wind-blown sediments produced in or near deserts. In northern China, nearly half a million square kilometres are covered by aeolian deposits⁷.

The well known loess–soil sequences of the last 2.6 million years (Myr) contain more than thirty major soil units interbedded with loess^{7,8}. These are underlain at some localities in the eastern Loess plateau by the Hipparion Red-Earth Formation (also called Red Clay)^{3,5,9–11}. Palaeomagnetic measurements date the lower boundary of the Red-Earth Formation to ~7–8 Myr ago^{9,10}. Detailed studies confirm that at least the portion of this sequence younger than 6.2 Myr is of aeolian origin⁵. This places the minimum age of initiation of sizeable deserts in Asia in the late Miocene^{3,5}.

Thick silty deposits with numerous palaeosols are also widely distributed in the western loess region, an area between the Liupan mountains and the Tibetan plateau, representing one-fifth of the area of the Loess plateau (Fig. 1). Neither the origin nor stratigraphy of these layers has been investigated in detail. The sequences reported here were found in Qinan county (Gansu province), about 160 km northeast of the Tibetan plateau and about 300 km south of the Tengger desert (Fig. 1). The present-day climate at Qinan is semi-arid with a mean annual rainfall of 400 mm and an annual mean temperature of 10.4 °C.

The region has experienced strong erosion, and its topography is characterized by valleys flanked by elongated hills with thick and relatively flat sedimentary sequences deposited on metamorphic bedrock of early Palaeozoic age. At the bottom of the sedimentary

sequences lies a waterlain deposit 10–30 m thick, consisting of reddish clay and containing abundant stratified sandy layers and bedrock fragments. Hill tops are unconformably mantled by pale-brown Quaternary-period loess up to 30 m thick. In between lies the Miocene loess sequence reported here.

The most complete loess sequence (QA-I) (105° 27' E, 35° 2' N) is 253.1 m in thickness and located 27 km northwest of Qinan city (Fig. 1), and a second section (QA-II, 220.6 m thick) is located 2 km from QA-I. QA-I contains 231 visually definable reddish layers interbedded with yellow-brown or brown silty layers, all nearly horizontal (Fig. 2). This layering is reflected in variations of magnetic susceptibility (Fig. 3), with higher values in red layers than in surrounding silty units. According to field lithostratigraphic correlation, the top part of QA-II was truncated by erosion. A spatially correlative erosion surface was observed at a depth of 59.8 m in QA-I and 20.0 m in QA-II.

The reddish layers have a silty clay or clay texture and moderate to strong angular blocky or prismatic structure, characteristic of soils¹². The abundant biological pores and clay coatings (Fig. 2) observed in microscopic thin sections confirm that these red beds are palaeosols, many of them composite soils. Fine-grained ferri-magnetic minerals of pedogenic origin may be responsible for their higher susceptibility values, as in Quaternary loess–soil sequences in China¹³.

Most soils are underlain by carbonate nodule horizons similar to those in Quaternary loess¹⁴. The non-palaeosol silty layers have a yellow-brown or brown colour and massive structure. Thin-section examination shows that the coarse fraction (>10 µm) is dominated by quartz (50–60%) and feldspar and mica (>30%). All mineral grains have angular shapes with no significant rounding (Fig. 2). The grain size is generally finer than 70 µm with the <20 µm fraction amounting to approximately 80%. The field aspects, mineral assemblage, grain morphology and grain-size distribution are characteristic of aeolian dust deposits^{6,7,15}. This is also confirmed

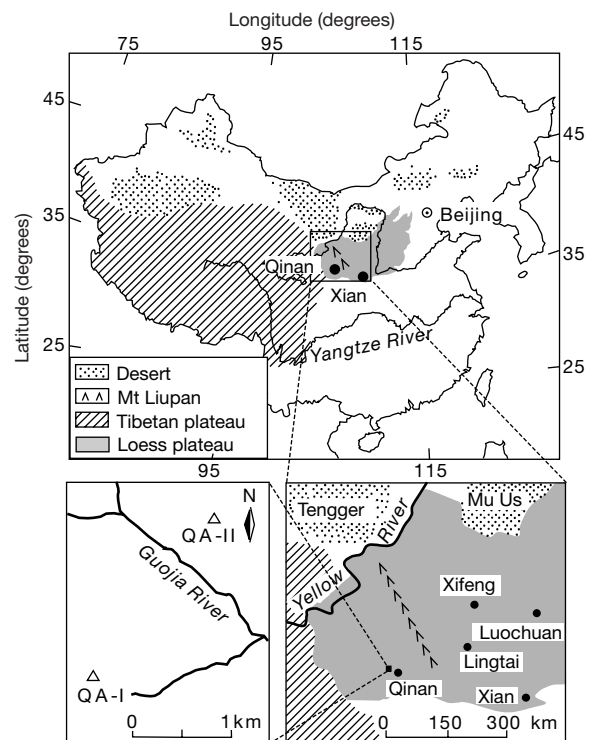


Figure 1 Maps showing the Loess plateau, Tibetan plateau, deserts in northern China and location of sites studied.

by the high degree of geochemical similarity to the Quaternary loess (Fig. 2).

Micromammalian fossils at Site QA-I were examined to determine the approximate age of the sequence. The seven fossil assemblages (Table 1) indicate ages from early Miocene to late Miocene^{16–20}. Palaeomagnetic measurements were made on both the QA-I and QA-II sections. After removing the secondary viscous remanent magnetization, clear polarity zonations were obtained (Fig. 3). The magnetozone of QA-I correlate with polarity intervals C6An.2n to C3An.2n in the geomagnetic polarity timescale (GPTS)²¹, and only C5n.1n and C5n.1r are missing from the sequence. The lack of these polarity intervals in QA-I, and of polarity intervals C4Ar.1r to C5n.1r in QA-II, is attributable to the erosion surfaces in both sections. This erosional break occurred around 9–10 Myr ago.

Extrapolation of the prevailing accumulation rate below the

oldest geomagnetic boundary yields a basal age of about 22 Myr for the QA-I section. The polarity zones below the erosion surface at a depth of 20.0 m in QA-II are correlative with C5n.2n to C6An.2n and give a basal age of about 21.6 Myr. The chronology obtained from magnetostratigraphy is consistent with the fossil assemblages (Table 1).

In summary, palaeomagnetic measurements date the main Qinan sections from around 22 to 6.2 Myr ago, except for hiatuses spanning around 0.7 Myr in QA-I and 1.2 Myr in QA-II. This gives an average dust accumulation rate at QA-I of approximately 1.67 cm kyr⁻¹. With 231 soils and aeolian layers visible in a sequence covering 15.1 Myr (after accounting for the hiatus), the average duration of each loess–soil pair in QA-I is 65,000 years, within the bandwidth typical of orbital-scale variations²². In summary, we have clear evidence that a baseline shift in Asian climate occurred by at least 22 Myr ago, and that aeolian sediments began to be deposited

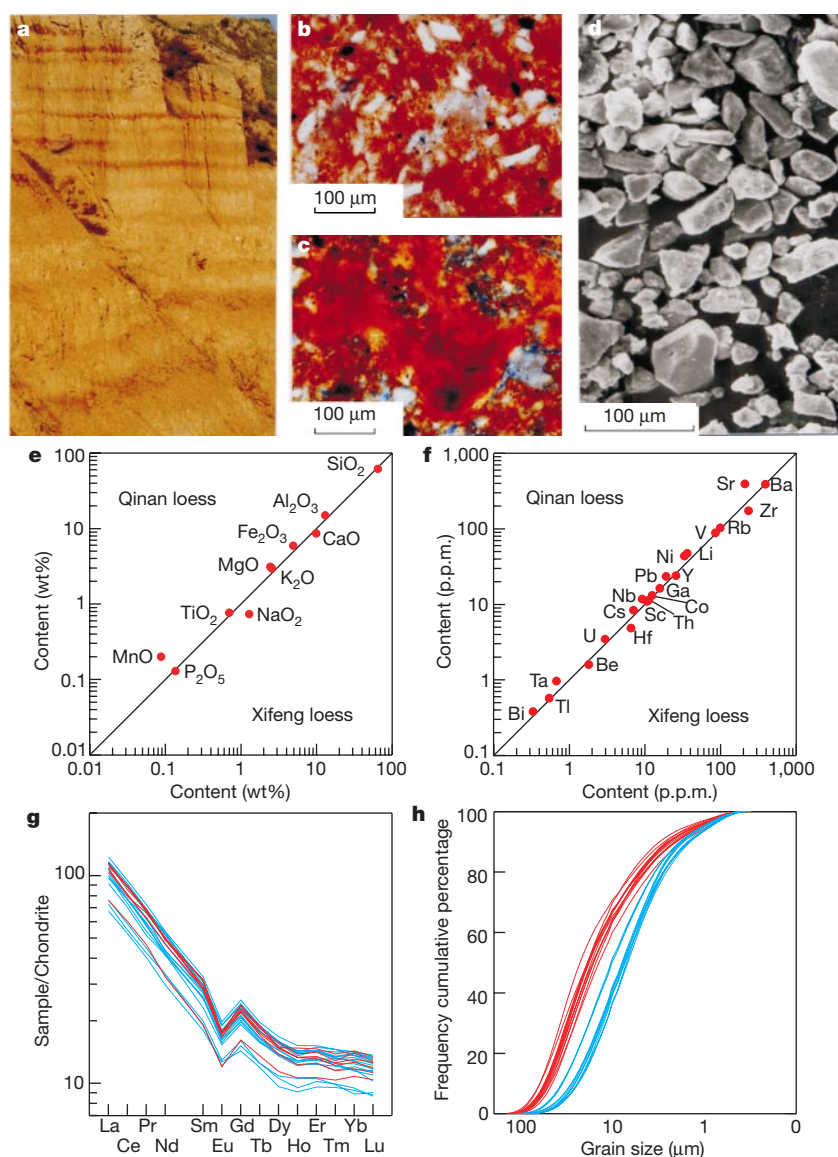


Figure 2 Morphological, geochemical and sedimentological properties of the Qinan aeolian deposits. **a**, Field picture showing the alternation of loess (yellow-brown layers) and palaeosols (reddish layers). **b**, Microscopic image showing the groundmass of a palaeosol. **c**, Microscopic image showing the clay coatings in palaeosols. **d**, Angular morphology of quartz grains under scanning electron microscope. **e**, **f**, Comparison of major and trace element compositions between the Qinan aeolian deposits (average of 16

samples from QA-I) and Quaternary loess from Xifeng (average of 6 samples).

g, Comparison of rare-earth elemental (REE) distribution patterns between the Qinan aeolian deposits (blue lines) and the Quaternary Xifeng loess (red lines). **h**, Grain-size distributions of the Qinan aeolian deposits (blue lines) and the Quaternary Xifeng loess (red lines).

on the loess plateau with orbital-scale cyclicity.

Loess deposits require both a sizeable source area arid enough to create aeolian particles and an atmospheric circulation sufficiently energetic to carry the particles from the deserts to the points of deposition. The high degree of geochemical similarity between the Miocene and the Quaternary loesses (Fig. 2) suggests similar sources for the dust. Quaternary dust deposits in northern China mainly originated from the desert lands in the north and northeast^{3,7,23}. By analogy, the Qinan aeolian sequences provide firm evidence that deserts large enough to produce aeolian dust must have been formed at or by 22 Myr ago in central Asia, some fourteen million years earlier than previously thought^{3,5,9–11}.

The Quaternary loess was probably carried from central Asia to the Loess plateau by northwesterly winds in the Asian winter monsoon. Today, strong northwesterly winds occur in the eastern portion of the high-pressure cell over Siberia^{3,7}. The Qinan deposits indicate that a winter wind pattern similar to the modern Asian winter monsoon had formed by the early Miocene. However, the finer texture of the Qinan loess compared with the Quaternary loess deposits (Fig. 2) suggests weaker wind strength and smaller extent of deserts. The alternations between loess and reddish soils indicate cyclical changes in intensity of winter and summer monsoons.

Three tectonic-scale changes, two regional in scope and one global, are probably causes of these joint changes in aridity and circulation. Interiors of large continents are normally arid because of the great distance to ocean moisture sources, but sediments deposited in the western Loess plateau of East Asia were fluvial-lacustrine through the Oligocene epoch²⁴. One source of Oligocene moisture in the Asian interior may have been the broad Paratethys sea in western-central Asia⁴. As this inland sea shrank during the late

Oligocene and early Miocene, it would have left the Asian interior more arid and would also have intensified the continentality, with hotter summers and colder winters. The onset of greater aridity recorded by the early Miocene loess is broadly consistent with this change in land–sea distribution.

A second regional-scale factor was uplift of the Tibetan–Himalayan complex, which began at smaller scales by 40 Myr ago²⁵, but intensified 25 to 20 Myr ago in the Himalayas, and southern Tibetan plateau²⁶. Geological evidence indicates rapid surface exposure and erosion of deeply buried rocks and minerals shortly after cooling. Because extremely rapid erosion only occurs in regions of high relief, this evidence indicates that sizeable parts of the south Tibetan–Himalayan complex were elevated by 25 to 20 Myr ago²⁶.

General circulation model experiments reveal the probable effects of large-scale Tibetan–Himalayan uplift on Asian climate^{1–3}. Uplift strengthens the summer monsoon and brings wetter climates to India and southeast Asia, but this moisture cannot reach the Asian interior because uplifted Himalayan topography blocks flow from the south. As a result, central Asia becomes drier as uplift proceeds. Uplift also produces drier climates in central Asia in the winter season because dry winter monsoon winds blow out of the Asian interior. The combination of summer and winter drying produces year-round aridity and forms deserts. In addition, uplift strengthens the flow of winter monsoon winds from the northwest. These model simulations^{1–3}, combined with geological evidence²⁶, provide an explanation for the early Miocene loess deposits: the southern margin of the Tibetan plateau was sufficiently elevated by 22 Myr ago to cause year-round drying and desert formation in the Asian interior and to produce northwest winds strong enough to carry aeolian particles southeast into the Loess plateau.

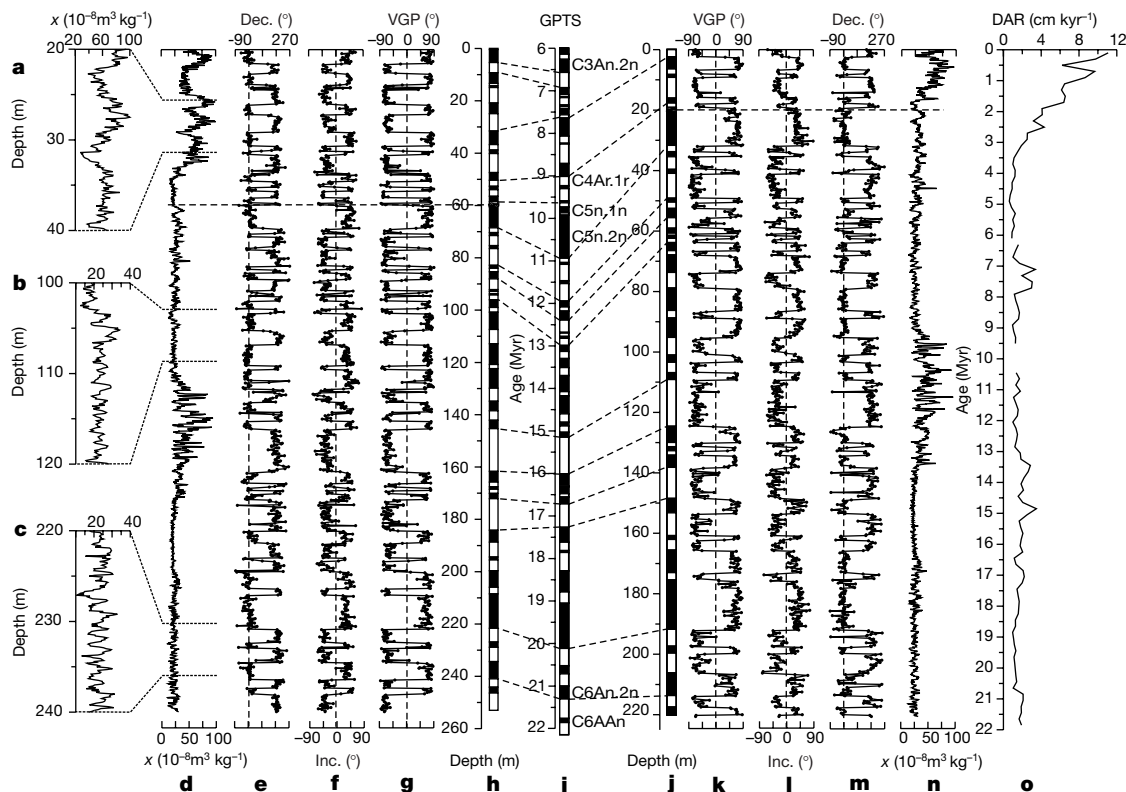


Figure 3 Magnetostratigraphy and magnetic susceptibility of the Qinan aeolian sequences and late Cenozoic dust accumulation rate in northern China. **a–h**, Profiles of magnetic susceptibility (χ), declination (Dec.), inclination (Inc.), virtual geomagnetic pole latitudes (VGP) and polarity zonation of QA-I. **i**, Reference geomagnetic polarity timescale (GPTS)²¹. **j–n**, Polarity zonations, VGP, Inc., Dec. and susceptibility profiles of QA-II. **o**, Dust accumulation rate (DAR) in northern China. Data for 0–6.2 Myr interval are

calculated based on the loess–soil sequence and the Hipparion Red-Earth formation at Xifeng. The 6.2–22.0 Myr portion is based on QA-II. Calculation was made at 200,000-year intervals. The timescales are obtained using geomagnetic boundaries (refs 8, 9 for Xifeng) as age controls and then interpolating, using the magnetic susceptibility age model for loess⁸. Horizontal discontinuous lines indicate the erosion surfaces in both sections.

Table 1 Micromammalian fossil assemblages from the QA-I section

Depth (m)	Assemblage	Fossils	Depth (m)	Assemblage	Fossils
8	A1	<i>Allopus cf. A. annectens</i> <i>Ochotona</i> sp. <i>Pararhizomys</i> sp. <i>Prosiphneus licenti</i>	197	A5	<i>Ansomys</i> sp. <i>Atlantoxerus</i> sp. <i>Sayimys minor</i> <i>Sinologomys cf. S. ulungurensis</i> <i>Sinologomys</i> sp. <i>Tachyoryctoididae</i> gen. et sp. indet.
36	A2	<i>Ochotona</i> sp. <i>Prosiphneus</i> n. sp.1	220	A6	<i>Distylomys cf. D. qianlishanensis</i> <i>Distylomys cf. D. tedfordi</i> <i>Litodonomys</i> sp. <i>Proditylomys cf. P. xinjiangensis</i> <i>Sinologomys cf. S. kansuensis</i> <i>Sinologomys cf. S. ulungurensis</i> <i>Ansomyinae</i> gen. et sp. indet.
82	A3	<i>Prosiphneus</i> n. sp.2 <i>Sciuridae</i> gen. et sp. indet.	227	A7	<i>Distylomys</i> sp. <i>Litodonomys</i> sp. <i>Tachyoryctoididae</i> gen. et sp. indet.
148	A4	<i>Alloptox gobiensis</i> <i>Ansomys</i> sp. <i>Cricetodon</i> sp. <i>Desmatolagus</i> sp. <i>Gobicricetodon</i> sp. <i>Mioechinus</i> sp. <i>Talpidae</i> gen. et sp. indet.			

The dominance of *Prosiphneus licenti* in A1 is characteristic of the late Miocene in China¹⁶ and was correlated to MN12 (8.2–7.1 Myr) in the European Neogene Mammal Chronology (ENMC)¹⁷. A4 is highly similar to the Tunggur fauna in China¹⁸, correlative with MN7–8 (12.5–11.2 Myr) in the ENMC. A5 is characteristic of the early Miocene in China¹⁹. *Sayimys minor* was geomagnetically dated in the Murree Formation in Pakistan to 18 Myr (ref. 20). A6 and A7 contain species that are abundant in the late Oligocene to early Miocene strata in northern China¹⁹.

Ongoing global cooling during the Cenozoic probably also contributed to regional changes in Asia. Although neither ice sheets nor extensive sea ice appeared in the Northern Hemisphere until late in the Miocene²⁷, climatic cooling had caused Palaeocene evergreen forests and Eocene-epoch deciduous forests along the Arctic margins to yield to mixed coniferous forests by the early Miocene²⁸. Oxygen-isotopic values from benthic foraminifers also indicate significant early Miocene cooling of polar oceans²⁹. Cooler high-latitude oceans provide less moisture to continental interiors, promoting continentality at the expense of temperate climates. Continental climates are also generally marked by colder, drier winters with stronger winds.

The Miocene sequences of Qinan do not show any obvious long-term intensification of loess deposition between 22 and 6.2 Myr. The major increases occur after about 3.5 Myr, in the Pliocene and Pleistocene epochs (Fig. 3). This suggests that the level of aridity and the strength of the winter monsoon winds in central Asia remained at moderate levels through the Miocene. Two intervals with higher dust accumulation rates are observed at 15–13 Myr and 8–7 Myr ago. The latter matches independent evidence reported from studies of North Pacific aeolian fractions³⁰. Because these intervals of higher aeolian deposition were not sustained, they may simply represent temporary instabilities of climate or land–surface conditions in the dust source region.

The early-Miocene climatic changes reported here intensified in the late Miocene and the Pliocene–Pleistocene, as part of the ongoing deterioration of Cenozoic climate. As sea ice expanded across the Arctic Ocean and ice sheets periodically formed in North America and northern Eurasia, the earlier Miocene trend towards a drier and windier climate in central Asia intensified (Fig. 3). Thick, well defined loess sequences eventually formed in the Loess plateau^{5,7,8}, replacing the thinner and finer-grained aeolian Miocene deposits reported here. □

Methods

Major chemical elements were analysed by X-ray fluorescence using a Philips PW-1400 unit with analytical uncertainties of 2% for all the major elements except for P₂O₅ and MnO (up to 10%). Trace and rare earth elements were digested with a HF + HClO₄ + HNO₃ mixture in Teflon bombs and analyses were made on a Finnigan MAT inductively coupled plasma mass spectrometry (ICP-MS) unit, which gave a precision of less than 10%. Grain-size distribution was determined using a Master Sizer 2000 laser unit.

Magnetic susceptibility was measured on air-dried samples as 10-cm intervals using a Bartington susceptibility meter. Palaeomagnetic measurements were made on 947 samples from QA-I and 846 samples from QA-II at 25–30-cm intervals in the Paleomagnetism Laboratory of the Institute of Geology and Geophysics, Chinese Academy of Sciences. Stepwise thermal demagnetization was performed on 661 samples from QA-I at an average interval of 40 cm. Samples were demagnetized in a MMTD-600 Thermal Demagnetiser and measured using a 2 G three-axis cryogenic magnetometer, both installed in field-free space. Because stepwise thermal demagnetization on 661 samples

from QA-I showed that most of the samples yield a stable characteristic remanent magnetization (ChRM) above 350 °C (see Supplementary Information), the other 286 samples from QA-I and all the samples from QA-II were then demagnetized at 350 and 400 °C. 874 (92%) samples from QA-I and 771 (91%) samples from QA-II gave reliable characteristic remanence directions. Details of the rock magnetic properties will be described elsewhere.

Received 27 September 2001; accepted 16 January 2002.

- Ruddiman, W. F. & Kutzbach, J. E. Forcing of late Cenozoic northern hemisphere climate by plateau uplift in southern Asia and the American West. *J. Geophys. Res.* **94**, 18409–18427 (1989).
- Manabe, S. & Broccoli, A. J. Mountains and arid climates of middle latitudes. *Science* **247**, 192–195 (1990).
- An, Z. S., Kutzbach, J. E., Prell, W. L. & Porter, S. C. Evolution of Asian monsoons and phased uplift of the Himalaya–Tibetan plateau since Late Miocene times. *Nature* **411**, 62–66 (2001).
- Ramstein, G., Fluteau, F., Besse, J. & Joussame, S. Effect of orogeny, plate motion and land-sea distribution on Eurasian climate change over the past 30 million years. *Nature* **386**, 788–795 (1997).
- Guo, Z. T., Peng, S. Z., Hao, Q. Z., Biscay, P. E. & Liu, T. S. Origin of the Miocene–Pliocene Red-Earth formation at Xifeng in northern China and implications for paleoenvironments. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **170**, 11–26 (2001).
- Whalley, W. B., Marshall, J. R. & Smith, B. J. Origin of desert loess from some experimental observations. *Nature* **300**, 433–435 (1982).
- Liu, T. S. *Loess and the Environment* (China Ocean, Beijing, 1985).
- Kukla, G., An, Z. S., Melice, J. L., Gavin, J. & Xiao, J. L. Magnetic susceptibility record of Chinese loess. *Trans. R. Soc. Edinb. Earth Sci.* **81**, 263–288 (1990).
- Sun, D. H., An, Z. S., Shaw, J., Bloemendal, J. & Sun, Y. B. Magnetostratigraphy and paleoclimatic significance of Late Tertiary aeolian sequences in the Chinese Loess Plateau. *Geophys. J. Int.* **134**, 207–212 (1998).
- Sun, D. H., Shaw, J., An, Z. S., Chen, M. Y. & Yue, L. P. Magnetostratigraphy and paleoclimatic interpretation of a continuous 7.2 Ma late Cenozoic eolian sediments from the Chinese Loess Plateau. *Geophys. Res. Lett.* **25**, 85–88 (1998).
- Ding, Z. L. *et al.* Wind-blown origin of the Pliocene red clay formation in the central Loess Plateau, China. *Earth Planet. Sci. Lett.* **161**, 135–143 (1998).
- Duchaufour, Ph. *Pédologie Tome 1 Pédogenèse et Classification* (Masson, Paris/New York/Barcelona/Milan, 1983).
- Maher, B. A. Characterization of soils by mineral magnetic measurements. *Phys. Earth Planet. Inter.* **42**, 76–92 (1986).
- Guo, Z. T. *et al.* Climate extremes in loess of China coupled with the strength of deep-water formation in the North Atlantic. *Glob. Planet. Change* **18**, 113–128 (1998).
- Pye, K. & Sperling, C. H. B. Experimental investigation of silt formation by static breakage processes: the effect of temperature, moisture and salt on quartz dune sand and granitic regolith. *Sedimentology* **30**, 49–62 (1983).
- Zheng, S. H. in *Rodent and Lagomorph Families of Asian Origins and Diversification* (eds Tomida, Y., Li, C. K. & Setoguchi, T.) 57–76 (National Science Museum Monographs 8, Tokyo, 1994).
- Mein, P. in *European Neogene Mammal Chronology* (eds Lindsay, E. H., Fahlbusch, V. & Mein, P.) 73–90 (NATO ASI Series A 180, Plenum, New York, 1990).
- Qiu, Z. D. *Middle Miocene Micromammalian Fauna from Tunggur, Nei Mongol* (Science, Beijing, 1996).
- Qiu, Z. X., Xu, W. Y. & Qiu, Z. D. in *The Miocene Land Mammals of Europe* (eds Rössner, G. & Heissig, K.) 443–455 (Pfeil, München, 1999).
- Brujin, H., Hussain, S. T. & Leinders, J. J. M. Fossil rodents from the Murree formation near Banda Daud Shah, Kohat, Pakistan. *Proc. Koninklijke Nederlandse Akad. Wetenschappen Ser. B* **84**, 71–99 (1981).
- Cande, S. C. & Kent, D. V. Revised calibration of the geomagnetic polarity timescale for the late Cretaceous and Cenozoic. *J. Geophys. Res.* **100**, 6093–6095 (1995).
- Berger, A., Loutre, M. F. & Laskar, J. Stability of the astronomical frequencies over the Earth's history for paleoclimate studies. *Science* **255**, 560–566 (1992).
- Zhang, X. Y., Arimoto, R. & An, Z. S. Dust emission from Chinese desert sources linked to variations in atmospheric circulation. *J. Geophys. Res.* **102**, 28041–28047 (1997).
- Wang, H. Z. (ed.) *Atlas of the Paleogeography of China* 121–123 (Cartographic, Beijing, 1985).

25. Chung, S. L. *et al.* Diachronous uplift of the Tibetan plateau starting 40 Myr ago. *Nature* **394**, 769–773 (1998).
26. Harrison, T. M., Copeland, P., Kidd, W. S. F. & Yin, A. Raising Tibet. *Science* **255**, 1663–1670 (1992).
27. Jansen, E. & Sjöholm, J. Reconstruction of glaciation over the past 6 Myr from ice-borne deposits in the Norwegian Sea. *Nature* **349**, 600–603 (1991).
28. Wolfe, J. A. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archaean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 357–375 (Monograph 32, American Geophysical Union, Washington DC, 1985).
29. Miller, K. G., Fairbanks, R. G. & Mountain, G. S. Tertiary oxygen isotope synthesis, sea level history, and continental margin erosion. *Paleoceanography* **2**, 1–19 (1987).
30. Rea, D. K., Snoeck, H. & Joseph, L. H. Late Cenozoic eolian deposition in the North Pacific: Asian drying, Tibetan uplift, and cooling of the northern hemisphere. *Paleoceanography* **13**, 215–224 (1998).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank W. Wu, Z. Qiu, S. Zheng and Z. Qiu for the identification of micromammalian fossils, and Z.S. An, J. Guiot and R. Potts for discussions. This work is supported by the National Natural Science Foundation of China, the National Project for Basic Research on Tibetan Plateau, Chuangxin and Bairen Programs of the Chinese Academy of Sciences.

Correspondence and requests for materials should be addressed to Z.T.G. (e-mail: ztguo@95777.com).

Strong emission of methyl chloride from tropical plants

Yoko Yokouchi*, Masumi Ikeda†‡, Yoko Inuzuka* & Tomohisa Yukawa‡

* National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki 305-8506, Japan

† Meisei University, 2-1-1 Hodokubo, Hino, Tokyo 191-8506, Japan

‡ Tsukuba Botanical Garden, National Science Museum, 4-1-1 Amakubo, Tsukuba, Ibaraki 305-0005, Japan

Methyl chloride is the largest natural source of ozone-depleting chlorine compounds, and accounts for about 15 per cent of the present atmospheric chlorine content¹. This contribution was likely to have been relatively greater in pre-industrial times², when additional anthropogenic sources—such as chlorofluorocarbons—were absent. Although it has been shown that there are large emissions of methyl chloride from coastal lands in the tropics^{3,4}, there remains a substantial shortfall in the overall methyl chloride budget. Here we present observations of large emissions of methyl chloride from some common tropical plants (certain types of ferns and Dipterocarpaceae), ranging from 0.1 to 3.7 µg per gram of dry leaf per hour. On the basis of these preliminary measurements, the methyl chloride flux from Dipterocarpaceae in southeast Asia alone is estimated at 0.91 Tg yr⁻¹, which could explain a large portion of missing methyl chloride sources. With continuing tropical deforestation, natural sources of chlorine compounds may accordingly decrease in the future. Conversely, the abundance of massive ferns in the Carboniferous period⁵ may have created an atmosphere rich in methyl chloride.

We studied CH₃Cl emissions from tropical plants in the lowland forest section (25 m × 20 m × 10–24 m high) of the Tropical Rainforest Glasshouse in Tsukuba Botanical Gardens, where more than 200 representative species from lowland tropical forests of southeast Asia grow. In order to see if tropical forests could be a significant source of CH₃Cl and to estimate the magnitude of the flux, CH₃Cl concentration in the glasshouse was measured over a few days in early September and mid October of 2000. During the experiment in October, air ventilation in the glasshouse was manually controlled to determine the accumulation of CH₃Cl. Air was sampled in the middle of the glasshouse from the observation catwalk (4 m

high), as well as from outside, upwind of the glasshouse, by opening pre-evacuated canisters. CH₃Cl and some other volatile organic compounds were analysed with pre-concentration/capillary gas chromatography/mass spectrometry after samples were transported to the laboratory. Details of the analytical method have been published elsewhere^{6,7}.

Figure 1 shows the concentrations of CH₃Cl as well as carbonyl sulphide (COS), which is taken up by plants⁸. Ventilation status during the experiments is shown at the bottom of Fig. 1, along with the timings of the intentional (non-automatic) ventilation when the indoor atmosphere was entirely replaced with fresh outside air. CH₃Cl concentrations were always higher in the glasshouse (filled circles) than outside (open circles), and increased significantly when the windows were closed. Similar variations, but with smaller differences in concentration, were also observed for CH₃Br and CH₃I, clearly showing the emission of all these methyl halides from this tropical rainforest system. On the other hand, COS concentration was lower in the glasshouse; its variation was in strong contrast to that of CH₃Cl. The fluxes of CH₃Cl and COS were calculated from the averages of their accumulation rates (Δp.p.t.v. h⁻¹) under closed conditions and the dimensions of the glasshouse (20 m × 25 m × 17 m average height), giving 5.4 (3.8–8) µg m⁻² h⁻¹ for CH₃Cl and –0.28 (–0.11 to –0.3) µg m⁻² h⁻¹ for COS.

As a first step toward global extrapolation, the above CH₃Cl emission rate per unit area, 5.4 µg m⁻² h⁻¹ (or 47 mg m⁻² yr⁻¹) was multiplied by the total area of tropical forest (1.73 × 10¹³ m²; ref. 9), giving an annual emission of 0.82 Tg yr⁻¹. This estimate is expected to suffer from substantial uncertainty owing to the lower density and lower diversity of the plants in the glasshouse compared to those in actual tropical forests. Another extrapolation was done using the ratio of CH₃Cl emission to COS uptake in the glasshouse (100:5.2). By using the reported COS uptake by tropical forests (0.43 ± 0.10 Tg yr⁻¹; ref. 8), we calculated an annual CH₃Cl flux of 8.2 Tg yr⁻¹, a figure that exceeded even the total known sinks (3.5 Tg yr⁻¹; ref. 1). This estimate would not be affected by the difference in the density of plants, but would be revised downward if the ratio of CH₃Cl emission to COS uptake for tropical forests were lower. In any case, we may say that tropical forests are an important source of CH₃Cl on a global scale.

In order to determine which of the plants was responsible for the increase of CH₃Cl, or whether the soil was responsible, we measured flux by using an enclosure method. We chose 19 families (see Table 1) of plants from among the 55 families in the glasshouse

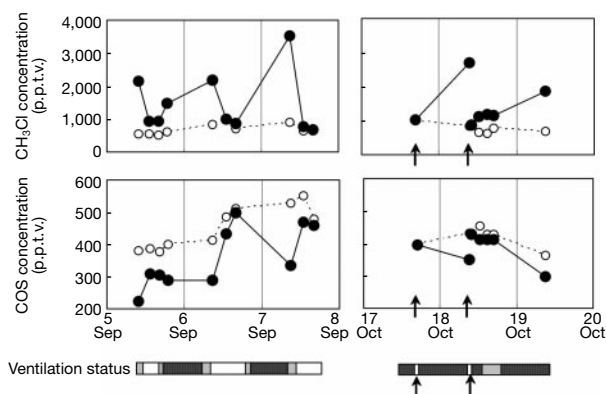


Figure 1 Atmospheric concentrations of CH₃Cl and COS during the observation periods, 5–8 September and 17–20 October. Filled circles, concentrations inside the Tropical Rainforest House; open circles, concentrations outside. Bar underneath time axis shows ventilation status: unshaded, windows mostly open; grey, windows partly open; black, windows mostly closed. Arrows under bar show times of intentional ventilation to replace the indoor atmosphere with fresh outside air. p.p.t.v., parts per trillion by volume.

by giving priority to those families contributing most to the biomass or variety of species in the glasshouse, and we measured the flux of one or two species from each family (we tested a total of 30 plants). We used a flux chamber (40 cm in diameter by 25 cm height) to measure gas fluxes from the soil, and Teflon bags (approximately 20 l) to measure fluxes from the plants. The CH_3Cl flux was calculated from the difference of the concentration in the air sample extracted 10 min after enclosure and in that collected near the target plant. For some plants, the flux was also calculated on the basis of two samples of air extracted with an interval of 10 min, giving the same results as above. The soil was found to take up CH_3Cl at a small rate (from 0.2 to $0.6 \mu\text{g m}^{-2} \text{h}^{-1}$). On the other hand, some plants from the Marattiaceae, Cyatheaceae (tree fern), Dicksoniaceae and Dipterocarpaceae families were found to emit CH_3Cl significantly. The first three families are ferns commonly growing in tropical forests, and Dipterocarpaceae species are dominant in the tropical rainforests of southeastern and southern Asia. The flux measurements were done in May/June and in October of 2001, and the results are summarized in Table 1a. We observed CH_3Cl emissions as high as $0.53\text{--}1.9 \mu\text{g per g dry leaf per h}$ for *Cibotium balometz* (Dicksoniaceae), and the average CH_3Cl emission rate from the nine plants in Table 1a was around $0.4 \mu\text{g per g dry leaf per h}$. Replicate measurements taken of *Angiopteris pruinosa* and *Shorea guiso* showed that the CH_3Cl emission rate from a plant was rather constant on a short timescale (days), not depending on the time of day. The total CH_3Cl fluxes from these nine plants (which were calculated from the emission rate per unit leaf weight in Table 1a and estimated total leaf weight for each plant) amounted to 2 mg h^{-1} in May/June and 1.3 mg h^{-1} in October, while the total emission rate in the glasshouse calculated from the accumulation rate in the glasshouse (Fig. 1) was 3.5 mg h^{-1} in early September and 1.9 mg h^{-1} in October. This suggests that a small group of plants belonging to Marattiaceae, Cyatheaceae, Dicksoniaceae and Dipterocarpaceae is responsible for a large part of the CH_3Cl emissions in the glasshouse. Chloride content in the glasshouse soil was as low as $5\text{--}8 \text{ mg per kg dry soil}$, and was not likely to be related to the high CH_3Cl emission from the plants.

Table 1 Fluxes of methyl chloride from tropical plants

	Flux ($\mu\text{g per g dry leaf per h}$)*	
	May–June	October
(a) Tropical Rainforest Glasshouse†		
(Dipterocarpaceae)		
<i>Shorea guiso</i> (Blanco) Blume	0.15 (0.01)‡	0.18 (0.07)§
<i>Shorea multiflora</i> (Burck) Symington	0.37	0.20
<i>Shorea roxburghii</i> G. Don	0.62	0.38
(Marattiaceae)		
<i>Angiopteris pruinosa</i> Kunze	0.68 (0.08)‡	0.57 (0.12)¶
<i>Angiopteris boninensis</i> Hieron.	0.20	0.07
<i>Angiopteris palmiformis</i> (Cav.) C. Chr.	0.05	0.19
<i>Angiopteris lygodilifolia</i> Rosenst.	0.62	0.21
(Dicksoniaceae)		
<i>Cibotium balometz</i> (L.) J. Sm.	1.90	0.53
(Cyatheaceae)		
<i>Cyathea ogurae</i> (Hayata) Domin	0.08	0.19
(b) Subtropical forests in Okinawa island		
	Site A, March	Site B, October
(Cyatheaceae)		
<i>Cyathea lepifera</i> E. Copel (no. 1, 3 m high)	1.10	
<i>Cyathea lepifera</i> E. Copel (no. 2, 1.5 m high)		0.20 (0.07)¶
<i>Cyathea lepifera</i> E. Copel (no. 3, 3 m high)		0.15 (0.04)§
<i>Cyathea lepifera</i> E. Copel (no. 4, 4 m high)		1.84 (0.05)¶
<i>Cyathea lepifera</i> E. Copel (no. 5, 5 m high)		3.68

Numbers in parenthesis are standard deviation of replicate measurements.

* Ratio of dry weight to fresh weight was ~ 0.2 for the ferns, and ~ 0.45 for Dipterocarpaceae.

† This table displays results for CH_3Cl -emitting plants from the 19 families of plants studied: Marattiaceae, Araceae, Moraceae, Asclepiadaceae, Aspleniaceae, Palmae, Cyatheaceae, Cycadaceae, Davalliaceae, Polypodiaceae, Dilleniaceae, Pteridaceae, Dicksoniaceae, Dipterocarpaceae, Dryopteridaceae, Euphorbiaceae, Sapotaceae, Hernandiaceae and Leguminosae.

‡ 2–3 measurements from the same part of a plant on different days.

§ 3 measurements (morning, mid-day, late afternoon) from the same part of a plant.

¶ 3 measurements (morning, mid-day, late afternoon) from different parts of the same plant.

¶ 2 measurements (mid-day, late afternoon) from different parts of the same plant.

We also conducted several flux measurements from *Cyathea lepifera* E. Copel at a subtropical mixed forest (site A) and at a *Cyathea* forest dominated by *Cyathea lepifera* E. Copel (site B) in Okinawa, in March and in late October, respectively. All the measurements were done under clear weather conditions. High CH_3Cl emission amounting to $0.15\text{--}3.7 \mu\text{g per g dry leaf per h}$ was observed for *Cyathea lepifera* E. Copel (Table 1b). Replicate measurements of the CH_3Cl emission from *Cyathea lepifera* E. Copel at site B showed no significant change from different parts of one plant, but a considerable change from different individual plants. The CH_3Cl emission rate from *Cyathea lepifera* no. 3 measured at 10:00, 14:00 and 17:00 local time was almost constant at around $0.15 \mu\text{g per g dry leaf per h}$, while the isoprene (a major biogenic hydrocarbon) emission changed greatly from $0.5 \mu\text{g per g dry leaf per h}$ (at 17:00) to $166 \mu\text{g per g dry leaf per h}$ (at 14:00). These observations supported our findings on the CH_3Cl emission from the plants in the tropical glasshouse.

Thus, it is clear that some particular fern families and Dipterocarpaceae emit significant amounts of CH_3Cl . These plant families are very common in tropical forests¹⁰, and Dipterocarpaceae in particular contribute to 30% of total basal area in typical lowland forests in southeast Asia^{11,12}. Assuming 50% of the tropical forest area of Asia ($2.8 \times 10^{12} \text{ m}^2$)⁹ to be lowland, the CH_3Cl emission from only Dipterocarpaceae in Asian tropical forests could be estimated to be 0.91 Tg yr^{-1} using the average emission rate from the three species of Dipterocarpaceae in Table 1 ($0.32 \mu\text{g per g dry leaf per h}$) and the leaf biomass reported for mature tropical lowland rainforest (0.776 kg m^{-2} ; ref. 13). Considering the large variability of the CH_3Cl emission among species of a family and among individual plants of a species (as was found for *Cyathea lepifera*), this estimation should suffer from a great uncertainty. As for the fern families in Table 1, their contribution to the CH_3Cl emission seems to be much smaller because of their small biomass compared to Dipterocarpaceae. However, it is very probable that the CH_3Cl emission from Dipterocarpaceae all over the world as well as the other CH_3Cl -emitting tropical plants would fill much of the gap between the total known sources, 1.7 Tg yr^{-1} (the ocean, 0.4 (ref. 14); biomass burning, 0.91 (ref. 15); salt marshes, 0.17 (ref. 16); wood rot fungi, 0.16 (ref. 17); and rice paddies, 0.006 (ref. 18), and the total known sinks, 3.5 Tg yr^{-1} (atmospheric reaction; ref. 3). Moreover, a tropical forest source of CH_3Cl is consistent with previous observations on atmospheric CH_3Cl : tropical forests occupy vast areas along the Equator, and CH_3Cl concentration has a latitudinal gradient—highest in the tropics—which is almost uniform in the Northern and Southern Hemispheres. Enhanced CH_3Cl concentrations have in fact been observed in tropical forests⁴ and in the marine air masses affected by tropical islands^{4,19}.

From the present finding that tropical plants are strong sources of atmospheric CH_3Cl , two questions arise. First, will continued destruction of tropical rainforests cause reduction of atmospheric CH_3Cl in the future, and second, was much more CH_3Cl emitted in the Carboniferous period (250–350 million years ago) when ferns prospered⁵ as a major group of plants, forming the dense ‘coal age’ forests under a warm and mild climate? Further investigation of the tropical forest source of CH_3Cl (such as its emission rates from a broad range of species, and factors affecting emission) is required to understand the past trends of this compound (and to predict future trends), and its potential effects on atmospheric chemistry—such as stratospheric ozone depletion. □

Received 3 July 2001; accepted 29 January 2002.

- Butler, J. H. Better budgets for methyl halides? *Nature* **403**, 260–261 (2000).
- Butler, J. H. et al. A record of atmospheric halocarbons during the twentieth century from polar firm air. *Nature* **399**, 749–755 (1999).
- Khalil, M. A. K. & Rasmussen, R. A. Atmospheric methyl chloride. *Atmos. Environ.* **33**, 1305–1321 (1999).
- Yokouchi, Y. et al. A strong source of methyl chloride to the atmosphere from tropical coastal land. *Nature* **403**, 295–298 (2000).

5. Berg, L. *Introductory Botany* (Saimders College Publishing, New York, 1997).
6. Li, H.-J., Yokouchi, Y. & Akimoto, H. Measurements of methyl halides in the marine atmosphere. *Atmos. Environ.* **33**, 1881–1887 (1999).
7. Yokouchi, Y. *et al.* Isoprene in the marine boundary layer (Southeast Asian Sea, Eastern Indian Ocean, Southern Ocean): Comparison with DMS and bromoform. *J. Geophys. Res. D* **104**, 8067–8076 (1999).
8. Watts, S. F. The mass budgets of carbonyl sulfide, dimethyl sulfide, carbon disulfide and hydrogen sulfide. *Atmos. Environ.* **34**, 761–779 (2000).
9. Food and Agriculture Organization of the United Nations *State of the World's Forests 1997* (UN Food and Agriculture Organization, Rome, 1997).
10. Takhtajan, A. *Floristic Regions of the World* (Univ. California Press, Berkeley, 1986).
11. Manokaran, N. in *Conservation, management and development of forest resources* (ed. See, L. S. *et al.*) 41–60 (Forest Research Institute Malaysia, Malaysia, 1998).
12. Aiba, S. & Kitayama, K. Structure, composition and species diversity in an altitude-substrate matrix of rain forest tree communities on Mount Kinabalu, Borneo. *Plant Ecol.* **140**, 139–157 (1999).
13. Yamakura, T., Hagihara, A., Sukardjo, S. & Ogawa, H. Aboveground biomass of tropical rain forest stands in Indonesian Borneo. *Vegetatio* **68**, 71–82 (1986).
14. Moore, R. M., Groszko, W. & Niven, S. J. Ocean–atmosphere exchange of methyl chloride: results from NW Atlantic and Pacific Ocean studies. *J. Geophys. Res.* **101**, 28529–28539 (1996).
15. Lobert, J. M., Keene, W. C., Logan, J. A. & Yevich, R. Global chlorine emissions from biomass burning: Reactive chlorine emission inventory. *J. Geophys. Res. D* **104**, 8373–8389 (1999).
16. Rhew, R. C., Miller, B. R. & Weiss, R. F. Natural methyl bromide and methyl chloride emissions from coastal salt marshes. *Nature* **403**, 292–295 (2000).
17. Watling, R. & Harper, D. B. Chloromethane production by wood-rotting fungi and an estimate of the global flux to the atmosphere. *Mycol. Res.* **102**, 769–787 (1998).
18. Redker, K. R. *et al.* Emissions of methyl halides and methane from rice paddies. *Science* **290**, 966–969 (2000).
19. Li, H.-J., Yokouchi, Y., Akimoto, H. & Narita, Y. Distribution of methyl chloride, methyl bromide, and methyl iodide in the marine boundary air over the western Pacific and southeastern Indian Ocean. *Geochim. J.* **35**, 137–144 (2001).

Acknowledgements

We thank T. Konishi for permission to use the Rainforest Glasshouse, Tsukuba Botanical Garden; S. Yamauchi for permission to use Yanbaru Subtropical Forest Park; P. Ciccioli for suggestions on the measurements of flux; T. Okuda, A. Takenaka, K. Kitayama and T. Ueda for comments and discussion; and T. Enomoto for help with sampling.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.Y. (e-mail: yokouchi@nies.go.jp).

A Jurassic mammal from South America

Oliver W. M. Rauhut*, Thomas Martin†, Edgardo Ortiz-Jaureguizar* & Pablo Puerta*

* Museo Paleontológico Egidio Feruglio, Fontana 140, 9100 Trelew, Argentina

† Institut für Geologische Wissenschaften, Fachrichtung Paläontologie, Freie Universität Berlin, Malteserstraße 74–100, 12249 Berlin, Germany

The Jurassic period is an important stage in early mammalian evolution, as it saw the first diversification of this group, leading to the stem lineages of monotremes and modern therian mammals¹. However, the fossil record of Jurassic mammals is extremely poor, particularly in the southern continents. Jurassic mammals from Gondwanaland are so far only known from Tanzania^{2,3} and Madagascar⁴, and from trackway evidence from Argentina⁵. Here we report a Jurassic mammal represented by a dentary, which is the first, to our knowledge, from South America. The tiny fossil from the Middle to Late Jurassic of Patagonia is a representative of the recently termed Australosphenida, a group of mammals from Gondwanaland that evolved tribosphenic molars convergently to the Northern Hemisphere Tribosphenida, and probably gave rise to the monotremes¹. Together with other mammalian evidence from the Southern Hemisphere^{2–4,6–8}, the discovery of this new mammal indicates that the Australosphenida had diversified and were widespread in Gondwanaland well

before the end of the Jurassic, and that mammalian faunas from the Southern Hemisphere already showed a marked distinction from their northern counterparts by the Middle to Late Jurassic. Mammalia Linnaeus, 1758 (sensu McKenna and Bell 1997, ref. 9) Holotheria Wible *et al.*, 1995 (ref. 10) (sensu McKenna and Bell 1997, ref. 9)

Australosphenida Luo *et al.*, 2001 (ref. 1)

Asfaltomylos patagonicus gen. et sp. nov.

Etymology. Genus name referring to the Cañadón Asfalto Formation and mylos (Greek), mill, after the grinding function of the well developed talonid; the species name, *patagonicus*, is termed after the origin, from Patagonia.

Holotype. Museo Paleontológico Egidio Feruglio (Trelew) collection (MPEF-PV) 1671. Left mandible with roots and crown fragments of the last three premolars and M₁–M₃ (Figs 1 and 2).

Horizon and locality. The specimen comes from the Queso Rallado locality, approximately 4 km west-northwest of the village of Cerro Condor, Chubut, Argentina. The material is derived from a silicified mudstone within a series of lacustrine mudstones and limestones. The sequence can be assigned to the Cañadón Asfalto Formation. The Cañadón Asfalto Formation, which is in the vicinity of Cerro Condor, has repeatedly been dated as Callovian–Oxfordian (latest Middle to earliest Late Jurassic)^{11,12}.

Diagnosis. *Asfaltomylos* is a small holotherian mammal with fully basined talonids, with hypoconid and hypoconulid at the molars (broken on M₁). Trigonids are lingually open, the trigonid angle of M₁ is about 130° (obtuse), and for M₂ and M₃ it is about 80° (acute). All preserved teeth are double-rooted, M₁–M₃ with faint lingual cingulid at the base of the paraconid. There is no distal metacristid present on the talonids. The tooth formula of the mandible is: ?I, ?C, 3+P, 3M. The dentary is slender with a gently rising coronoid process. The angular process is posteroventrally positioned and the dental foramen is placed far towards the anterior (below the origin

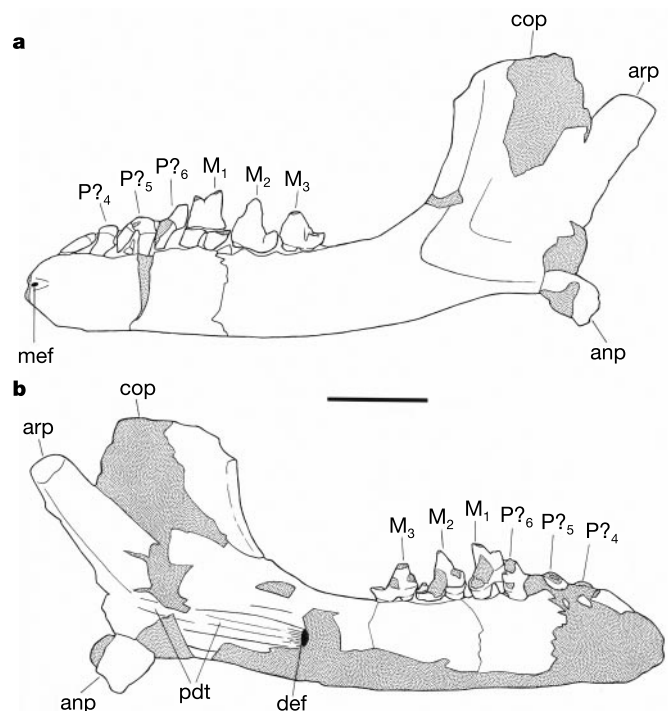


Figure 1 Camera lucida drawings of the left mandible of *Asfaltomylos patagonicus* (holotype). **a**, **b**, Labial (**a**) and lingual (**b**) view. Damaged areas are shaded. Scale bar, 2 mm. A small mental foramen is situated at the broken anterior end of the mandible. The numeration of the premolars follows comparisons with *Bishops*⁸. anp, angular process; arp, articular process; cop, coronoid process; def, dental foramen; M, molar; mef, mental foramen; P, premolar; pdt, post-dentary trough.

of the coronoid process). There is no developed pterygoid fossa. A shallow post-dentary trough is present, subdivided by longitudinal striations. Meckel's groove is not detectable.

Asfaltomylos differs from *Peramus* and other pre-tribosphenic holotherians through the presence of a broad, fully basined talonid with wear on the hypoconid and hypoconulid, lack of a distal metacristid, an anterior position of the dental foramen, presence of a post-dentary trough, and lack of pterygoid fossa. Differs from *Theria*, and *Holoclemensia* and *Pappotherium* by the presence of a lingually open trigonid. Important characters shared with australosphenidans are the presence of a lingual cingulid at the base of the paraconid, and talonids, which have a greater width than length. Differs from *Ambondro*, from the Middle Jurassic of Madagascar, by the lack of a distal metacristid and a weaker lingual cingulid, which does not wrap around the mesial side. The Early Cretaceous australosphenidan *Ausktribosphenos*, from Australia, is characterized by a number of distinctive autapomorphic features, such as

additional cristids on the metaconid and talonid¹³, which are not present in *Asfaltomylos*. Differs from the Early Cretaceous *Bishops*⁸ through the lack of additional cristids in the talonid, a weaker lingual cingulid, and a more slender protoconid.

Asfaltomylos is the first Jurassic mammal from South America and the fifth Mesozoic representative of the recently termed Australosphenida^{1,14}. Similar to *Ambondro* from the Middle Jurassic (Bathonian) of Madagascar⁴, it exhibits broad, fully rimmed, basined talonids on the molars (only the base of the talonid is preserved on M₁). On M₂ of *Asfaltomylos*, the talonid represents the broadest part of the tooth. A labially inflated, large hypoconid with pointed tip is present; it bears a lingually oriented wear facet. On M₃, hypoconid and hypoconulid are well developed and are connected by the talonid rim. On the lingual side, a part of the talonid rim is missing; just distally of the broken area the rim of the talonid rises, which possibly indicates the presence of a small entoconid. A large wear facet connecting hypoconid and hypoconulid is clearly visible.

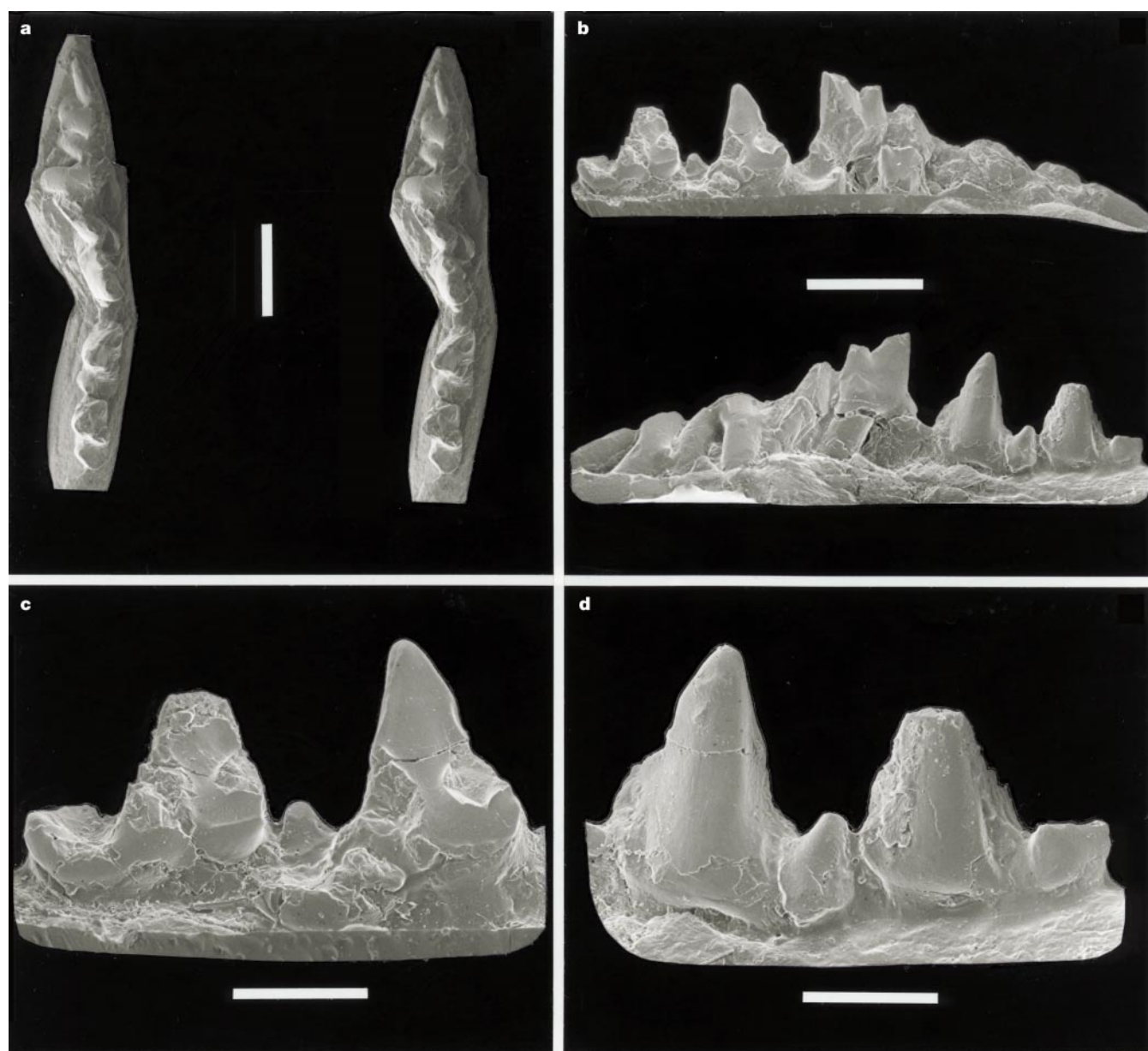


Figure 2 Dentition of the holotype of *Asfaltomylos patagonicus*. **a**, Occlusal view (stereopair); **b**, lingual (upper) and labial view (lower); **c**, detail of M₂ and M₃ in oblique lingual view; **d**, detail of M₂ and M₃ in oblique labial view. Scale bars: **a**, **b**, 1 mm;

c, **d**, 0.5 mm. M₁ has an obtuse angular trigonid; M₂ and M₃ have acute angular trigonids and fully basined talonids, which have a greater width than length. There is a faint lingual cingulid at the base of the paraconids of M₁ to M₃.

From this wear facet the presence of a functional protocone on the upper molar is inferred, making *Asfaltomylos* a fully tribosphenic mammal. High-resolution X-ray analysis did not reveal any trace of an unerupted tooth germ in the wide space behind M_3 .

Other important similarities to *Ambondro* include the basis of the paraconid being higher than that of the metaconid, and the talonid having a greater width than length. These characters constitute differences compared with the situation found in the Laurasian tribosphenic mammals, the boreosphenidans¹; together with the presence of a medial cingulid in the molars, they represent synapomorphies of the Australosphenida^{1,14}, thus supporting this concept and the placement of *Asfaltomylos* within this group.

The slender dentary of *Asfaltomylos* is broken anteriorly of P_4 and is crushed in the premolar region. On the partially broken lingual side, no Meckel's groove is visible (present in *Ausktribosphenos* from Australia^{7,13}). The subdivision of the post-dentary trough indicates the presence of distinct post-dentary bones¹⁵. This is a very primitive character¹⁶ that supports a basal phylogenetic position for australosphenidans^{1,14,17} within holotherians, despite their highly derived molar morphology. The angular process is well developed. As preserved, it is pointing ventrally

owing to compression, and the original orientation was most probably pointing posteriorly, as in other australosphenidans¹⁴. The articular process rises steeply and points posterodorsally. Below the origin of the coronoid process, opens the large dental foramen, which is at a very anterior position compared with *Peramus* and tribosphenic Zatheria^{18,19}.

Discussion. *Asfaltomylos* extends the fossil record of Jurassic mammals to South America, and corroborates the existence of a Middle Jurassic (or even earlier) radiation of tribosphenic holotherians on Gondwanaland. The development of tribosphenic molars on the southern continents predates that on Laurasia by about 25 Myr.

To test the phylogenetic position of *Asfaltomylos* within mammals, character codings for this taxon were added to the three data matrices published in the Supplementary Information of ref. 1 (see Supplementary Information). When we included *Asfaltomylos* in the phylogenetic analysis of ref. 1, it appeared as the most basal member of Australosphenida (Fig. 3).

The Australosphenida group to date is based on only a few taxa, and its history is still largely unknown. The oldest known representative is *Ambondro* from the Bathonian of Madagascar⁴. Only three other Mesozoic members of this clade are currently known, *Ausktribosphenos*, *Bishops* and *Steropodon*, all from the Early Cretaceous of Australia^{6–8,14}. The addition of *Asfaltomylos* to the analysis resolves the polytomy within Australosphenida found by ref. 1, and thus demonstrates that *Ambondro* and *Ausktribosphenos* represent successively closer outgroups to Monotremata, which include *Steropodon*. This pattern is in general accordance with the stratigraphic occurrence of these taxa.

Together with the occurrence of *Ambondro* in the Bathonian of Madagascar, the discovery of *Asfaltomylos patagonicus* in the Middle to Late Jurassic of southern South America indicates that the Australosphenida had already diversified and spread over much of Gondwanaland by the Middle Jurassic. Of note, no australosphenidan mammal has been reported from the Cretaceous period of South America, although a variety of mammalian taxa are known from that time. The South American mammalian fauna during the Cretaceous was dominated by Dryolestida, with a few 'triconodonts', gondwanatherians, and a questionable docodont^{20–23}, which may indicate that the Australosphenida had already largely been replaced on this continent by that time. Together with the gondwanatherians—which are known from the Cretaceous of Madagascar (*Lavanify*²⁰), India (undetermined taxon) and southern South America (*Ferugliotherium*, *Gondwanatherium*, *Sudamerica*^{20,24})—the surviving Australosphenida from the Cretaceous of Australia (*Ausktribosphenida*^{7,8}, *Steropodon*⁶) and the Palaeocene epoch of South America (*Monotrematum*^{25,26}) might represent relics of a Jurassic radiation of purely Gondwanan mammals. More fossil evidence, particularly from the Cretaceous of Africa and the Late Jurassic to Early Cretaceous of South America, is needed to test this hypothesis.

A second aspect worth noting is the already marked differentiation between the mammalian faunas of the Southern and Northern Hemispheres towards the end of the Jurassic. The late Middle to Late Jurassic mammalian faunas of Laurasia are dominated by multituberculates, symmetrodonts, docodonts, dryolestids, and stem-lineage representatives of modern therian mammals, such as peramurids^{27–30}. In contrast, the sparse fossil evidence from the Southern Hemisphere indicates a mammalian fauna consisting of endemic and highly derived groups, the Australosphenida, and very probably the Gondwanatheria²⁰, in addition to relics of an archaic group, the Haramiyida³, which is unknown from rocks younger than the Early Jurassic in the Northern Hemisphere. However, other taxa are still found on both landmasses, such as triconodonts, dryolestids, and peramurids², indicating that some faunal interchange might still have occurred towards the end of the Jurassic. Commonly, the great success of eutherian (and to some extent metatherian) mammals is attributed to the evolution of tribo-

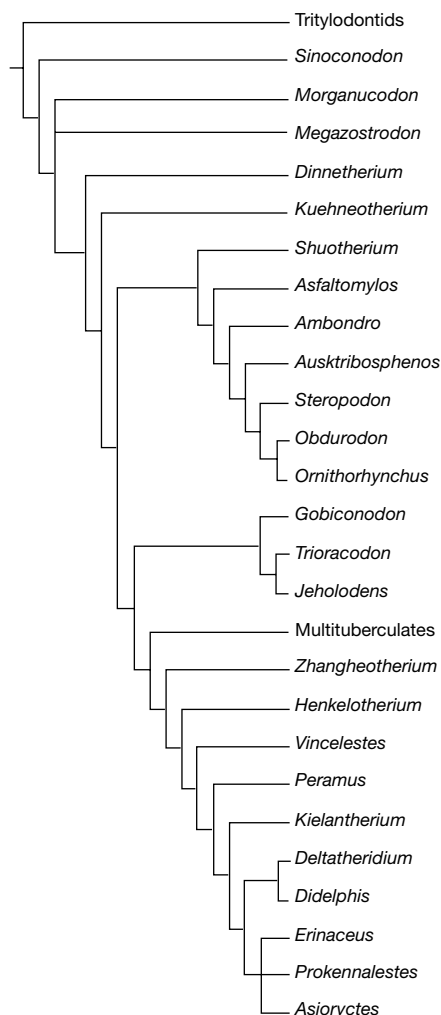


Figure 3 Phylogenetic position of *Asfaltomylos*. Inclusion of *Asfaltomylos* in a phylogenetic analysis of the main taxa of Mesozoic and several younger mammals (based on 125 characters from the whole skeleton using PAUP 3.1.1, with the heuristic branch-and-bound search, set at random with 100 replicas; see ref. 1 and Supplementary Information) markedly reduced the number of equally parsimonious trees and resolved the polytomy of *Ambondro*, *Steropodon* and *Ausktribosphenos*. *Asfaltomylos* appears as the most basal member of Australosphenida.

sphenic molars, which allowed a much more effective processing of food and energy exploitation. However, this was only the case in Boreosphenida (tribosphenic mammals from the Northern Hemisphere), whereas the Australosphenida apparently decreased in diversity since the Late Jurassic, and finally evolved completely reduced dentition, as seen in extant Australian monotremes (egg-laying mammals). □

Received 4 October; accepted 19 December 2001.

- Luo, Z., Cifelli, R. L. & Kielan-Jaworowska, Z. Dual origin of tribosphenic mammals. *Nature* **409**, 53–57 (2001).
- Heinrich, W.-D. Late Jurassic mammals from Tendaguru, Tanzania, East Africa. *J. Mamm. Evol.* **5**, 269–290 (1998).
- Heinrich, W.-D. First haramiyid (Mammalia, Allotheria) from the Mesozoic of Gondwana. *Mitt. Mus. Naturk. Berl., Geowiss.* **2**, 159–170 (1999).
- Flynn, J. J., Parrish, J. M., Rakotosamimanana, B., Simpson, W. F. & Wyss, A. R. A Middle Jurassic mammal from Madagascar. *Nature* **401**, 57–60 (1999).
- Casamiquela, R. M. Sobre la presencia de un mamífero en el primer elenco (icnológico) de vertebrados del Jurásico de la Patagonia. *Physis* **22**, 225–233 (1961).
- Archer, M., Flannery, T. F., Ritchie, A. & Molnar, R. First Mesozoic mammal from Australia—an Early Cretaceous monotreme. *Nature* **318**, 363–366 (1985).
- Rich, T. H. *et al.* A tribosphenic mammal from the Mesozoic of Australia. *Science* **278**, 1438–1442 (1997).
- Rich, T. H. *et al.* A second tribosphenic mammal from the Mesozoic of Australia. *Rec. Queen Vict. Mus.* **110**, 1–10 (2001).
- McKenna, M. C. & Bell, S. *Classification of Mammals Above the Species Level* (Columbia Univ. Press, New York, 1997).
- Wible, J. R., Rougier, G. W., Novacek, M. J., McKenna, M. C. & Dashzeveg, D. D. A mammalian petrosal from the Early Cretaceous of Mongolia: implications for the evolution of the ear region and mammalian relationships. *Am. Mus. Novit.* **3149**, 1–19 (1995).
- Tasch, P. & Volkheimer, W. Jurassic conchostracans from Patagonia. *Univ. Kansas, Paleont. Contrib.* **50**, 1–23 (1970).
- Musacchio, E. A., Beros, C. & Pujana, I. in *Bioestratigrafía de los Sistemas Regionales del Jurásico y el Cretácico de América del Sur* Vol. 1 (ed. Volkheimer, W.) 355–383 (CRICYT, Mendoza, 1986).
- Rich, T. H. *et al.* Early Cretaceous mammals from Flat Rocks, Victoria, Australia. *Rec. Queen Vict. Mus.* **106**, 1–35 (1999).
- Luo, Z.-X., Kielan-Jaworowska, Z. & Cifelli, R. L. In quest for a phylogeny of early mammals. *Acta Palaeont. Pol.* **47**, 1–78 (2002).
- Kermack, K. A., Mussett, F. & Rigney, H. W. Lower jaw of *Morganucodon*. *Zool. J. Linn. Soc.* **53**, 87–175 (1973).
- Luo, Z.-X. in *In the Shadow of the Dinosaurs—Early Mesozoic Tetrapods* (eds Fraser, N. C. & Sues, H.-D.) 98–128 (Cambridge Univ. Press, Cambridge, 1994).
- Kielan-Jaworowska, Z., Cifelli, R. L. & Luo, Z.-X. Alleged Cretaceous placentals from down under. *Lethaia* **31**, 267–268 (1998).
- Clemens, W. A. & Mills, J. R. E. Review of *Peramus tenuirostris* OWEN (Eupantotheria, Mammalia). *Bull. Brit. Mus. (Nat. Hist.) Geol.* **20**, 89–113 (1971).
- Kielan-Jaworowska, Z. & Dashzeveg, D. Eutherian mammals from the Early Cretaceous of Mongolia. *Zool. Scripta* **18**, 347–355 (1989).
- Krause, D. W., Prasad, G. V. R., Koenigswald, W. von, Sahni, A. & Grine, F. E. Cosmopolitanism among Late Cretaceous mammals. *Nature* **390**, 504–507 (1997).
- Bonaparte, J. F. Approach to the significance of the Late Cretaceous mammals of South America. *Berliner Geowiss. Abh.* **13**, 31–44 (1994).
- Bonaparte, J. F. Cretaceous tetrapods of Argentina. *Münchener Geowiss. Abh.* **30**, 73–130 (1996).
- Pascual, R., Goin, F. J., González, P., Ardolino, A. & Puerta, P. F. A highly derived docodont from the Patagonian Late Cretaceous: evolutionary implications for Gondwanan mammals. *Geodiversitas* **22**, 395–414 (2000).
- Pascual, R., Goin, F. J., Krause, D. W., Ortiz Jaureguizar, E. & Carlini, A. A. The first gnathic remains of *Sudamerica*: implications for gondwanan relationships. *J. Vert. Paleont.* **12**, 373–382 (1999).
- Pascual, R. *et al.* The first discovery of monotremes in South America. *Nature* **356**, 704–705 (1992).
- Pascual, R. *et al.* in *Platytypus and Echidnas* (ed. Augee, M. L.) 1–14 (The Royal Society of New South Wales, Sydney, 1992).
- Simpson, G. G. *A Catalogue of the Mesozoic Mammalia in the Geological Department of the British Museum* (British Museum of Natural History, London, 1928).
- Hahn, G. Beiträge zur Fauna der Grube Guimarota Nr. 3. Die Multituberculata. *Palaeontographica A* **133**, 1–100.
- Martin, T. Dryolestidae (Dryolestidae, Mammalia) aus dem Oberen Jura von Portugal. *Abh. Senckenberg. Naturforsch. Ges.* **550**, 1–119.
- Martin, T. & Krebs, B. (eds) *Guimarota. A Jurassic Ecosystem* (Friedrich Pfeil, Munich, 2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank E. Eggert for preparation of the specimen, M. Bulang-Lörcher for drawing Fig. 1, U. Gloy for electronic figures and W. Müller for assisting at the scanning electron microscope (SEM). We also thank J. Habersetzer for carrying out the high-resolution X-ray analysis. The manuscript benefitted from comments by Z. Kielan-Jaworowska. This work was supported by a Deutscher Akademischer Austauschdienst (DAAD) fellowship to O.R., a Deutsche Forschungsgemeinschaft (DFG) Heisenberg fellowship to T.M., and an Antorchas grant to the MPEF.

Correspondence and requests for materials should be addressed to T.M. (e-mail: tmartin@zedat.fu-berlin.de).

Bacterial growth and primary production along a north–south transect of the Atlantic Ocean

Hans-Georg Hoppe*, Klaus Gocke*, Regine Koppe* & Christian Begler†

Sections of * Marine Ecology and † Ocean Circulation and Climate, Institute of Marine Science, Düsterbrook Weg 20, D-24105 Kiel, Germany

The oceanic carbon cycle is mainly determined by the combined activities of bacteria and phytoplankton^{1,2}, but the interdependence of climate, the carbon cycle and the microbes is not well understood. To elucidate this interdependence, we performed high-frequency sampling of sea water along a north–south transect of the Atlantic Ocean. Here we report that the interaction of bacteria and phytoplankton is closely related to the meridional profile of water temperature, a variable directly dependent on climate. Water temperature was positively correlated with the ratio of bacterial production to primary production, and, more strongly, with the ratio of bacterial carbon demand to primary production. In warm latitudes (25° N to 30° S), we observed alternating patches of predominantly heterotrophic and autotrophic community metabolism. The calculated regression lines (for data north and south of the Equator) between temperature and the ratio of bacterial production to primary production give a maximum value for this ratio of 40% in the oligotrophic equatorial regions. Taking into account a bacterial growth efficiency^{3,4} of 30%, the resulting area of net heterotrophy (where the bacterial carbon demand for growth plus respiration exceeds phytoplankton carbon fixation^{4–6}) expands from 8° N (27 °C) to 20° S (23 °C). This suggests an output of CO₂ from parts of the ocean to the atmosphere^{6,7}.

We report here large-scale distributions of microbial variables as determined from the north–south Atlantic transect of the JGOFS (Joint Global Ocean Flux Study) *Polarstern* expedition ANT X (53° N to 65° S) during November 1991 to January 1992 (Fig. 1). High-frequency water sampling was performed continuously at intervals of 4–8 h, corresponding to 50–100 nautical miles, and summing up to 170 samples along the transect. Water samples were taken at 11 m depth from the moving ship.

The measured chlorophyll distribution along the transect corresponded well to the time-integrated pattern of chlorophyll recorded by the CZCS (coastal zone colour scanner) satellite (Fig. 1a). The meridional profiles of chlorophyll and bacterial production (in terms of tritiated leucine incorporation (TL_i; Fig. 1b) were similar. In detail, however, chlorophyll concentrations showed a wider overall range than bacterial production, and bacterial production in relation to chlorophyll was high—particularly in the oligotrophic tropical regions⁸ (Fig. 1a and b). Owing to seasonal and regional (extremely high microbial values on the Patagonian shelf) effects, higher mean values and ranges of microbial variables were measured in the south than in the north (Table 1).

In contrast to the individual variables, the ratio (expressed in per cent) of bacterial production (TL_i) to primary production was highly correlated with temperature in the North as well as in the South Atlantic (Table 2). The ratio between bacterial production and primary production changed significantly from about 2–10% in the cold and temperate climate zones to 40% in the tropics (Fig. 2). This indicates a large-scale temperature-dependent shift from net-autotrophic to increasingly heterotrophic conditions in the surface of the Atlantic. The strongly scattered individual data points around the regression lines suggest an alternating predominance of autotrophic and heterotrophic regimes in patches along the transect (Fig. 2). The estimates derived from tritiated thymidine

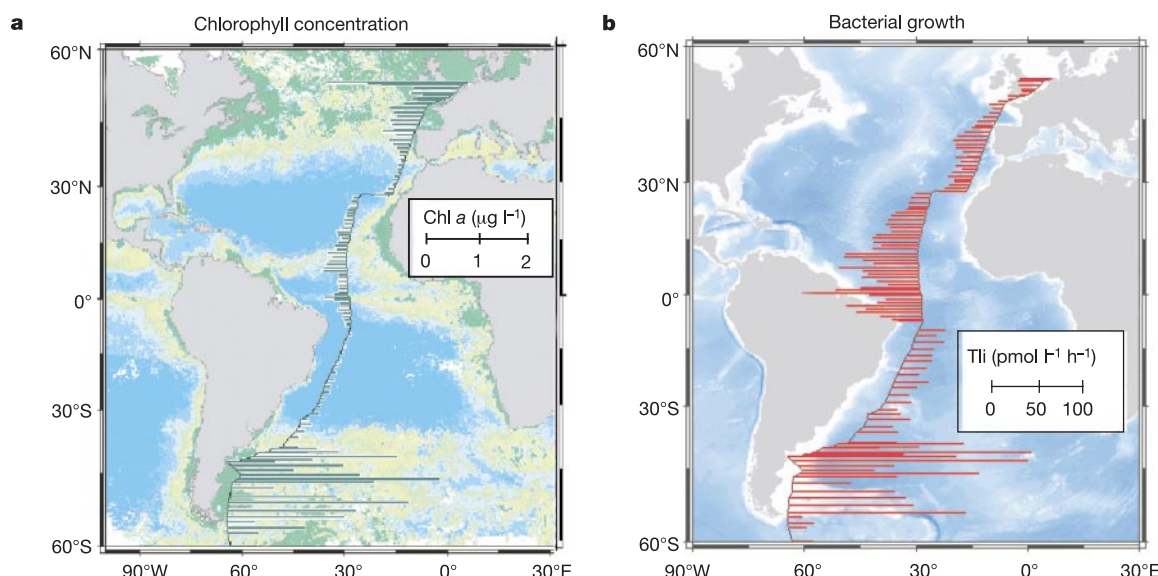


Figure 1 Profiles of chlorophyll concentration (**a**) and bacterial growth (**b**) along the Atlantic transect. Bacterial growth was determined by tritiated leucine incorporation (TL_i). The background of **a** is given by satellite (CZCS) images of chlorophyll distribution

integrated over several years. The background of **b** shows the depth distribution of the Atlantic Ocean.

incorporation (TT_i)—which represent only the bacterial production by reproducing cells—were also strongly correlated with the meridional temperature profile. The different slopes of the regression lines derived from the two methods indicate a shift from balanced to unbalanced bacterial growth towards the Equator.

The determination of net-heterotrophic areas on the meridional scale (Fig. 3) is based on the observed correlation between temperature and the ratio of bacterial versus primary production (Fig. 2). 'Net heterotrophy' means in this case that phototrophic carbon-fixation is lower than bacterial carbon-demand (BCD) for production plus respiration, regardless of zooplankton respiration. For the calculation of BCD the application of a factor for bacterial growth efficiency (BGE) is necessary. Excluding extreme values,

BGE is in the range of <10% to 50% (refs 4 and 9). Using a medium BGE of 30% and a conservative factor of 3H -leucine conversion to bacterial carbon production (see Methods), the resulting net-heterotrophic area extends from $8^\circ N$ ($27^\circ C$) to $20^\circ S$ ($23^\circ C$) (Fig. 3). Along the transect, 19 of the 90 sampling stations were net heterotrophic, with a clear predominance of net heterotrophy in the warm oligotrophic regions¹⁰. In addition, taking into account the variability of BGEs in relation to temperature, BGE was calculated according to the equation of ref. 11 (see Methods). In this case the resulting net-heterotrophic area extends from $22^\circ N$ ($25^\circ C$) to $32^\circ S$ ($17^\circ C$), and 30 of the 90 sampling stations were net heterotrophic (Fig. 3).

Our calculations of net-heterotrophic areas are based on a large

Table 1 Comparison of base parameters from the North and South Atlantic

	Temperature	Chlorophyll	Primary productions, PP	3H leucine incorporation, TL_i	TL_i/PP	3H -thymidine incorporation, TT_i	Protease activity
North Atlantic ($0^\circ N$ – $53^\circ N$)							
Mean	20.77	0.35	0.72	33.32	17.64	1.22	0.89
s.d.	5.48	0.43	0.65	22.81	20.02	0.79	0.43
Range	9.7–27.8	0.04–2.8	0.04–3.0	6.52–134	1.06–83	0.29–3.9	0.3–1.79
South Atlantic ($0^\circ S$ – $63^\circ S$)							
Mean	14.57	0.86	1.31	62.39	23.17	4.26	3.26
s.d.	10.10	0.95	2.10	58.81	28.19	5.39	3.02
Range	–1.22–26.6	0.04–3.9	0.03–11.0	12.6–264	2.27–119	0.55–21.4	0.82–10.9

The units of the variables are temperature, $^\circ C$; chlorophyll, $\mu g l^{-1}$; PP, $\mu g C l^{-1} h^{-1}$; bacterial growth TL_i , $pmol l^{-1} h^{-1}$; bacterial reproduction TT_i , $pmol l^{-1} h^{-1}$; protease activity $\mu g C l^{-1} d^{-1}$; TL_i/PP , %. For this calculation values of TL_i were converted to bacterial C production by a conservative factor of 1.5 kg C per mol of leucine. s.d., standard deviation.

Table 2 Correlation matrix of variables from the North and South Atlantic

	Temperature	TL_i	PP	TL_i/PP BGE 30%	TL_i/PP BGE variable	TT_i	Protease/ TL_i
North Atlantic ($53^\circ N$ – $0^\circ N$)							
Latitude	–0.983‡	–0.731‡	0.414	–0.711‡	–0.830‡	–0.503‡	0.686‡
Temperature		0.707‡	–0.436*	0.715‡	0.837‡	0.478‡	–0.659‡
South Atlantic ($0^\circ S$ – $63^\circ S$)							
Latitude	0.986‡	0.253	–0.549‡	0.782‡	0.871‡	–0.230	–0.630‡
Temperature		0.241	–0.549‡	0.787‡	0.882‡	0.233	–0.625‡

Spearman rank correlation coefficients of variables in the North and South Atlantic. Depending on n the indicated P values are: * $P < 0.01$, † $P < 0.001$, ‡ $P < 0.0001$. 'BEG variable' means that bacterial growth efficiency was calculated according to ref. 11. the positive or negative correlations obtained with latitude or temperature needs to be interpreted with regard to the orientation of these independent variables north and south of the Equator. Abbreviations are defined in Table 1.

number of horizontal samples instead of depth-integrated profiles. We attempted to compensate for this by submitting the samples to a medium light intensity (see Methods). Along the transect this would approximately fit the reduced-light conditions in the moderate climate zones. With respect to the bright (sub)-tropical regions it would mean an acceleration to the reduced-light conditions in the subsurface maxima of primary production. Nevertheless, the deep maxima of primary production were not sampled and thus, particularly in the bright regions, our values of primary production may not reflect depth-integrated values. For correction we have consulted the satellite-derived values of integrated primary production provided by Longhurst¹². During the months of observation, our values of primary production were 16% lower than Longhurst's values in his "Eastern Subtropical Province" and 14% or 17% lower in his "Tropical Gyral Province" and "Western Tropical Province", respectively. There was a greater discrepancy in the "South Atlantic Gyral Province" (43%), which may be due to the northward shift of the subtropical convergence during austral spring. If our values underestimated primary production by 17%, this would result in a minor shrinkage of net-heterotrophic areas for about 5°. Concerning the ratio of bacterial production to primary production, our values of 1–9% (using a conservative factor of leucine conversion) are at the lower end of the 5–15% reported from the temperate North Atlantic¹³. The same holds true for the Southern Ocean, where 12–22% were measured¹⁴ compared to 4–11% measured by us.

The correlation matrix of variables (Table 2) shows that bacterial production (TL_i) was more closely related to temperature in the North than in the South Atlantic. In contrast, primary production was highly correlated with temperature in the South, but less significantly in the North Atlantic. Only the ratio of bacterial production (TL_i , combined with a constant or variable factor of BGE) versus primary production was most significantly correlated with temperature in both global regions. In addition, the ratio between protease activity and bacterial production decreased

significantly from the high latitudes towards the Equator. This indicates a greater need of substrate hydrolysis for bacterial supply in the cold/temperate regions, while the availability of exudates ready for uptake possibly reduced the importance of enzymatic hydrolysis in the warm regions.

Microbes have great potential for predicting the effect of climate change at the biological level because they are mainly responsible for the evolution and assimilation of greenhouse gases. Large- and medium-scale distributions of microbial variables in oceanic regions have frequently been reviewed^{2,9,15,16}, modelled^{17,18} and reported from cruises^{14,19,20}. Temperature regulation of microbial variables has been discussed by several authors^{11,19,21} and an increase of bacterial respiration—relative to production—with temperature has been found¹¹. Our study demonstrates now a significant correlation, particularly for the ratio of bacterial versus primary production and the meridional temperature profile in the Atlantic Ocean. This is not surprising, because both of the contributing microbial variables are influenced by temperature in combination with other factors (light, nutrients, exudation, substrates). But temperature dependence is the common factor. Nutrient and substrate availability or grazing may dominate the regulation of microbial activities in restricted areas and periods of time but—at the meridional scale—the effects of these factors are overlaid by the regulatory capacity of temperature. However, high temperatures often coincide with oligotrophic conditions. Thus the correlation between temperature and the microbial variables may be influenced by the trophic situation.

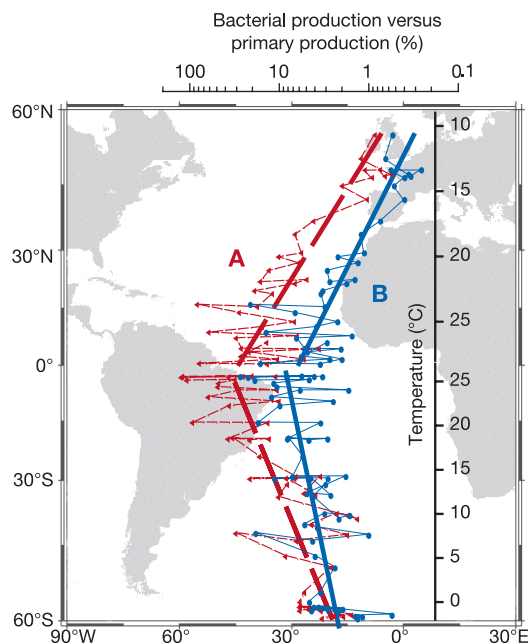


Figure 2 Ratio of bacterial production to primary production (PP) in (%) on the meridional temperature scale. Bacterial production was determined by tritiated leucine incorporation (TL_i ; A, dashed red lines) or by tritiated thymidine incorporation (TT_i ; B, solid blue lines). Regression A, north: slope $s = 0.079$, $r^2 = 0.67$, $n = 37$; south: $s = 0.036$, $r^2 = 0.58$, $n = 53$. Regression B, north: $s = 0.064$, $r^2 = 0.68$, $n = 37$; south: $s = 0.018$, $r^2 = 0.33$, $n = 53$. Note the scattering of the individual data points of A, approaching 100% (the criterion of net heterotrophy; see text) particularly in the warm regions of the Atlantic.

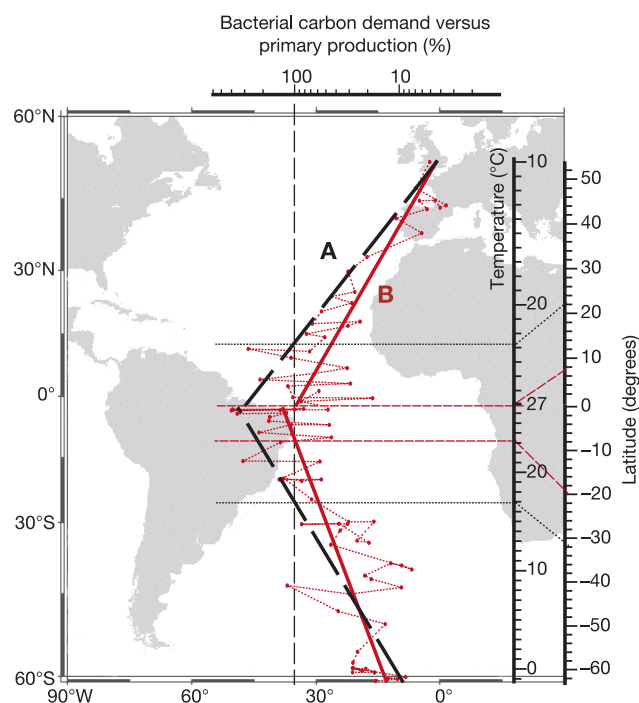


Figure 3 Net heterotrophy in the Atlantic Ocean. The estimation of net-heterotrophic areas (where bacterial carbon demand, BCD, exceeds primary production, PP) depends on conversion factors applied to the basic data. Factors used for the calculation of BCD were: A, Bold dashed regression lines (black), 1.5 kg C per mol of leucine and BGE calculated from the regression¹¹ $BGE = 0.3774 - 0.0104T$. The equations of the regression lines are: $\log[BCD/PP (\%)] = -0.427 + 0.107T$, $r^2 = 0.803$ (north) and $\log[BCD/PP (\%)] = 1.038 + 0.056T$, $r^2 = 0.763$ (south). B, Bold solid regression lines (red), 1.5 kg C per mol of leucine and 30% BGE. The equations are: $\log[BCD/PP (\%)] = -0.143 + 0.079T$, $r^2 = 0.675$ (north) and $\log[BCD/PP (\%)] = 1.166 + 0.036T$, $r^2 = 0.584$ (south). The extensions of net-heterotrophic areas as a function of temperature or latitude are indicated by the dashed and dotted lines. The data points refer to regression B.

The ocean-wide pattern of 'net heterotrophy' is essential for understanding the global carbon balance. The occurrence of net-heterotrophic domains in the equatorial ocean was hypothesized by Sorokin⁵ in 1971. In the current debate, net heterotrophy has been claimed^{4,6,10} or questioned^{22–24} by several authors. The co-variation of temperature and microbial variables observed in this study (Figs 2 and 3) requires an additional supply by carbon sources in the net-heterotrophic areas. Such supplies are derived from: (1) equatorial upwelling providing aged dissolved organic carbon, (2) strong currents (North and South Equatorial Currents, Counter Currents) transporting materials from remote upwelling zones into the area (Canary, Benguela), big rivers (Amazon, Congo, Niger) and, possibly, (3) the temporal transfer of organic matter as a slow-to-degrade fraction of primary production⁷. The combined effects of these additional carbon sources support the assumption that net heterotrophy in parts of the (sub)tropical Atlantic is possible. However, seasonal variability in bacterial production covering two orders of magnitude even in the oligotrophic Atlantic (J. M. Gasol, personal communication) together with diel variation²⁵ suggest that this may not be a permanent phenomenon and that the location of net-heterotrophic patches may be variable.

If external carbon sources were not sufficient to support net heterotrophy^{24,26}, storage of organic matter within the 'microbial loop' may buffer the mismatch between bacterial carbon demand and primary production—at least temporarily. Thus, net heterotrophy in the oligotrophic ocean may partly be influenced by characteristics of the 'microbial loop'. The proximity of heterotrophic and autotrophic patches in the oligotrophic ocean suggests an unstable system reacting rapidly to changing environmental conditions. Nevertheless—on a large scale—a trend towards net heterotrophy in the warm regions can be determined. □

Methods

General

Methods were standardized following the international JGOFS protocols. Water samples were taken by a special pump. Measurements obtained from samples taken with Niskin bottles were similar to those of pump samples. Incubation for microbial activity measurements was 2–3 h at *in situ* temperature. Generally 3 replicates and 1 blank were prepared. The standard deviation of replicates for all radiotracer experiments was generally $\pm 5\%$ of the mean value.

Phytoplankton variables

Chlorophyll was measured fluorometrically after ethanol extraction. Primary production was measured every 90–100 nautical miles under constant light conditions, at $150 \mu\text{E m}^{-2} \text{s}^{-1}$ for 8 h. Two replicates and one blank (dark incubation) were performed. Primary production was calculated in terms of $\mu\text{g C l}^{-1} \text{h}^{-1}$, for only the hours of daylight. To harvest even the very small primary producers in the tropical regions, 0.2- μm membrane filters were used for filtration. In the case of 0.2- μm filtration, fractions of the phytoplanktonic exudation products which were consumed by bacteria during the period of incubation may be co-harvested. Exudation can be $<5\%$ to $>30\%$ of the measured primary production, with high values particularly under oligotrophic conditions. Adding exudation to the values of primary production will shift the regression lines (Figs 2 and 3) to lower positions, which means a shrinkage of the net-heterotrophic areas.

Bacterial production

Bacterial production^{27,28} was measured via ^3H -leucine incorporation and ^3H -thymidine incorporation. Final concentrations of 53 nM of leucine (3 nM ^3H -leucine plus 50 nM unlabelled leucine) or 5.79 nM of ^3H -thymidine were used. For the conversion of leucine incorporation to bacterial production, a conservative factor of 1.55 kg C per mol of leucine was applied. This factor is based on that of ref. 28, disregarding isotope dilution. Thymidine incorporation was converted by a factor of 13.75 kg C per mol of thymidine²⁷.

Bacterial carbon demand

Estimation of bacterial carbon demand (BCD) is based on bacterial production measured via ^3H -leucine incorporation. ^3H -leucine incorporation data were preferred to the data of ^3H -thymidine incorporation because they include the production of organic matter by both cellular growth and reproduction of cells. A constant factor of growth efficiency (BGE) of 30% and variable factors computed from the equation of Rivkin and Legendre¹¹ $\text{BGE} = 0.3774 - 0.0104T$, where T is temperature, were used for the calculation of BCD.

Bacterial enzyme activity

The rate of bacterial protein hydrolysis was determined by the fluorogenic substrate

analogue leucine-MCA (leucine-methylcoumarinylamid) at concentrations ranging from 0.1 to 100 μM (ref. 29). Incubation of water samples in cuvettes was for 4–12 h, corresponding to water temperature. Fluorescence resulting from substrate hydrolysis was measured in a Kontron SM 25 fluorometer at 364 nm excitation and 445 nm emission wavelengths.

Statistical analysis

Statistical analysis of the data sets was performed by Sigma-plot and Sigma-stat analytical tools (test of normal distribution, least-square fit method, non-random Spearman rank distribution). Normal distribution of the dependent variables was usually obtained after log-transformation of the experimental data.

Received 27 April; accepted 14 December 2001.

1. Azam, F. Microbial control of oceanic carbon flux: The plot thickens. *Science* **280**, 694–696 (1998).
2. Ducklow, H. W. The bacterial component of the oceanic euphotic zone. *FEMS Microb. Ecol.* **30**, 1–10 (1999).
3. Amon, R. M. W. & Benner, R. Rapid cycling of high-molecular-weight dissolved organic matter in the ocean. *Nature* **369**, 549–552 (1994).
4. del Giorgio, P. A., Cole, J. J. & Cimleris, A. Respiration rates of bacteria exceed phytoplankton production in unproductive aquatic systems. *Nature* **385**, 148–151 (1997).
5. Sorokin, Y. I. Bacterial populations as components of oceanic ecosystems. *Mar. Biol.* **11**, 101–105 (1971).
6. Duarte, C. M. & Agusti, S. The CO_2 balance of unproductive aquatic ecosystems. *Science* **281**, 234–236 (1998).
7. Azam, F., Steward, G. F. & Ducklow, H. W. Significance of bacteria in carbon fluxes in the Arabian Sea. *Proc. Ind. Acad. Sci.* **103**, 341–351 (1994).
8. Simon, M., Cho, B. C. & Azam, F. Significance of bacterial biomass in lakes and the ocean: comparison to phytoplankton biomass and biogeochemical implications. *Mar. Ecol. Prog. Ser.* **86**, 103–110 (1992).
9. Azam, F. et al. The ecological role of water-column microbes in the sea. *Mar. Ecol. Prog. Ser.* **10**, 257–263 (1983).
10. Duarte, C. M., Agusti, S., Aristegui, J., Gonzales, N. & Anadon, R. Evidence for a heterotrophic subtropical northeast Atlantic. *Limnol. Oceanogr.* **46**, 425–428 (2001).
11. Rivkin, R. B. & Legendre, L. Biogenic carbon cycling in the upper ocean: Effects of microbial respiration. *Science* **291**, 2398–2400 (2001).
12. Longhurst, A. *Ecological Geography of the Ocean* (Academic, London, 1998).
13. Zubkov, M. V., Sleight, M. A. & Burkill, P. H. Heterotrophic bacterial turnover along the 20°W meridian between 59°N and 37°N in July 1996. *Deep-Sea Res. II* **48**, 987–1001 (2001).
14. Lochte, K., Bjornsen, P. K., Giesenhausen, H. & Weber, A. Bacterial standing stock and production and their relation to phytoplankton in the Southern Ocean. *Deep-Sea Res. II* **44**, 321–340 (1997).
15. Bird, D. F. & Kalf, J. Empirical relationship between bacterial abundance and chlorophyll concentration in fresh and marine waters. *Can. J. Fish. Aquat. Sci.* **41**, 1015–1023 (1984).
16. Cole, J. J., Pace, M. L. & Findlay, S. Bacterial production in fresh and saltwater ecosystems: a cross-system overview. *Mar. Ecol. Prog. Ser.* **43**, 1–10 (1988).
17. Hurtt, G. C. & Armstrong, R. A. A pelagic ecosystem model calibrated with BATS data. *Deep-Sea Res. II* **43**, 653–683 (1996).
18. Fasham, M. J. R., Boyd, P. W. & Savidge, G. Modelling the relative contributions of autotrophs and heterotrophs to carbon flow at a Lagrangian JGOFS station in the Northeast Atlantic. *Limnol. Oceanogr.* **44**, 80–94 (1999).
19. Kirchman, D. L., Rich, J. H. & Barber, R. T. Biomass and biomass production of heterotrophic bacteria along 140°W in the equatorial Pacific: effect of temperature on the microbial loop. *Deep-Sea Res. II* **42**, 603–619 (1995).
20. Karrasch, B., Hoppe, H.-G., Ullrich, S. & Podewski, S. The role of mesoscale hydrography on microbial dynamics in the northeast Atlantic: Results of a spring bloom experiment. *J. Mar. Res.* **54**, 99–122 (1996).
21. Rivkin, R. B., Anderson, M. R. & Lajzerowicz, C. Microbial processes in cold oceans. 1. Relationship between temperature and bacterial growth rate. *Aquat. Microb. Ecol.* **10**, 243–254 (1996).
22. Geider, R. J. Photosynthesis or planktonic respiration? *Nature* **388**, 132 (1997).
23. Williams, P. J. le B. The balance of plankton respiration and photosynthesis in the open oceans. *Nature* **394**, 55–57 (1998).
24. Williams, P. J. le B. & Bower, D. G. Regional carbon imbalances in the oceans. *Science* **284**, 1735 (1999).
25. Kuipers, B., van Noort, G. J., Vosjan, J. & Herndl, G. J. Diel periodicity of bacterioplankton in the euphotic zone of the subtropical Atlantic Ocean. *Mar. Ecol. Prog. Ser.* **201**, 13–25 (2000).
26. Raymond, P. A. & Bauer, J. E. Riverine export of aged terrestrial organic matter to the North Atlantic Ocean. *Nature* **409**, 497–500 (2001).
27. Fuhrman, J. A. & Azam, F. Thymidine incorporation as a measure of heterotrophic bacterioplankton production in marine surface waters: Evaluation and field results. *Mar. Biol.* **66**, 109–120 (1982).
28. Simon, M. & Azam, F. Protein content and protein synthesis rates of planktonic marine bacteria. *Mar. Ecol. Prog. Ser.* **51**, 201–213 (1989).
29. Hoppe, H.-G. in *Current Methods in Aquatic Microbial Ecology* (eds Kemp, P. F., Sherr, B. F., Sherr, E. B. & Cole, J. J.) 423–431 (CRC, Boca Raton, 1993).

Acknowledgements

We thank K. Lochte and K. Jürgens for discussions, and Rory P. Wilson for linguistic corrections. We thank DFG and BMBF for support of our JGOFS programmes.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.-G. H. (e-mail: hhoppe@ifm.uni-kiel.de).

Hearing visual motion in depth

Norimichi Kitagawa* & Shigeru Ichihara*

* Department of Psychology, Tokyo Metropolitan University, 1-1, Minami-Osawa, Hachioji-shi, Tokyo 192-0397, Japan

Auditory spatial perception is strongly affected by visual cues. For example, if auditory and visual stimuli are presented synchronously but from different positions, the auditory event is mislocated towards the locus of the visual stimulus—the ventriloquism effect^{1,2}. This ‘visual capture’ also occurs in motion perception in which a static auditory stimulus appears to move with the visual moving object^{3,4}. We investigated how the human perceptual system coordinates complementary inputs from auditory and visual senses. Here we show that an auditory aftereffect occurs from adaptation to visual motion in depth. After a few minutes of viewing a square moving in depth, a steady sound was perceived as changing loudness in the opposite direction. Adaptation to a combination of auditory and visual stimuli changing in a compatible direction increased the aftereffect and the effect of visual adaptation almost disappeared when the directions were opposite. On the other hand, listening to a sound changing in intensity did not affect the visual changing-size aftereffect. The results provide psychophysical evidence that, for processing of motion in depth, the auditory system responds to both auditory changing intensity and visual motion in depth.

The visual-motion aftereffect is a well documented phenomenon in which, after a prolonged viewing of a moving visual pattern in a particular direction, a stationary pattern appears to move in the opposite direction^{5,6}. Similar auditory aftereffects have been reported for such stimuli as horizontal motion^{7,8}, spectral motion⁹, intensity change^{10,11} and frequency change¹². It has been implicitly

assumed that motion aftereffects reflect fairly low-level neural processing and that they do not occur across sensory modalities, that is, adaptation in one modality produces aftereffects in the same modality only. Here we combine two aftereffects in different sensory modalities: an auditory changing-loudness aftereffect^{10,11} and a visual changing-size aftereffect¹³.

In our first experiment, we investigated whether adaptation to visual size-changing affects the auditory aftereffect. The image of an object changing in size produces a strong sensation that the object is moving in depth¹⁴. The visual adapting stimulus of 2 s duration was a white square either expanding or contracting between 0 and 2 deg at $\pm 1 \text{ deg s}^{-1}$. It appeared to move towards the observers from far away or vice versa. Sinusoids of 1,000 Hz were used as auditory stimuli for adaptation and test. The auditory adapting stimulus was either increased or decreased in level between 20 and 60 dB sound pressure level at $\pm 20 \text{ dB s}^{-1}$. The duration was also 2 s. Thus, the auditory and visual adapting stimuli were perfectly synchronized. The duration of the test tone was 1.5 s and the sound pressure level of the onset was set at 40 dB. We measured the magnitude of the auditory aftereffect after adaptation to each of the eight conditions, two adapting directions (increasing or decreasing) $\times$ four combinations of auditory and visual adapting stimuli: (1) auditory stimulus only (A+, A-); (2) auditory and visual stimuli changing in a compatible direction (A+V+, A-V-); (3) auditory and visual stimuli changing in an opposite direction (A+V-, A-V+); and (4) visual stimulus only (V+, V-). In the adapting conditions, one of the combinations of adapting stimuli was presented 60 times, after which only the auditory test stimulus was presented and the observers judged the direction of loudness change, not the direction of motion in depth.

The results (Fig. 1) showed that adaptation to visual size-changing influenced the auditory changing-loudness aftereffect. Ehrenstein and Reinhardt-Rutland¹⁵ found slight shifts in auditory localization after adaptation to visual horizontal motion. However, we found a strong auditory aftereffect after adaptation to visual stimulus alone (V+, V-). After viewing the square changing in size, the subjects perceived a steady test tone as though it changed in loudness in the direction opposite to that of the square's change in

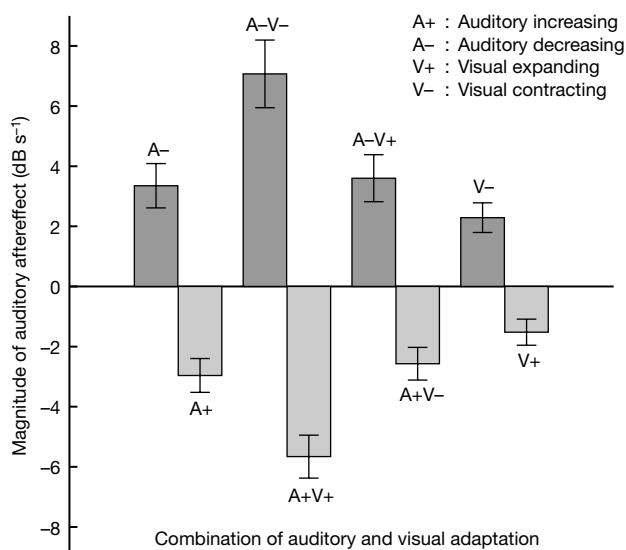


Figure 1 Magnitude of the auditory changing-loudness aftereffect after adaptation to combinations of auditory changing-intensity and visual changing-size stimuli. Averages of the ten subjects are shown. Error bars indicate the standard error of the mean. All of the four combination conditions produced the significant aftereffect ($P < 0.01$). Two-way analysis of variance (ANOVA) (adapting direction $\times$ adapting combination) for absolute value of the magnitude showed a significant main effect of the adapting combination ($F_{3,27} = 15.204$, $P < 0.01$), whereas the main effect of direction is not significant. A posterior pairwise comparison (Tukey's Honestly Significant Difference, HSD) of the adapting combination showed that the aftereffect of the same-direction combination (A+V+, A-V-) was stronger than the other combination conditions ($P < 0.01$).

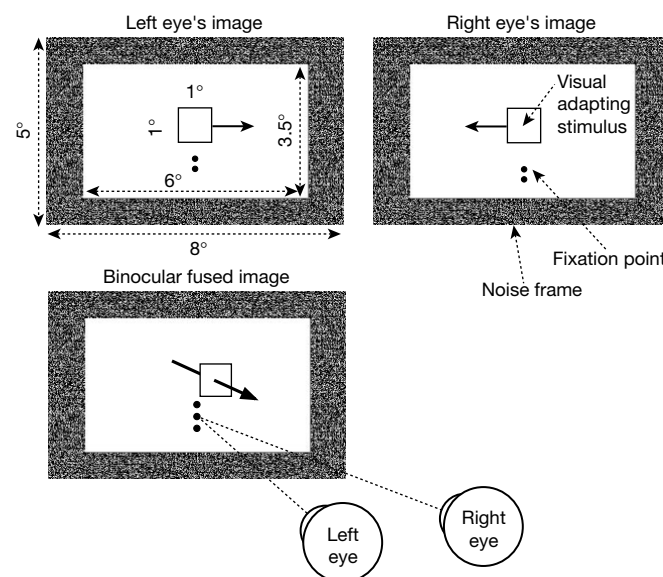


Figure 2 Schematic illustration of approaching visual stimulus. Each pattern for left and right eyes consisted of a white square and a frame of noise pattern surrounding the square. The patterns viewed by the left and right eyes were stationary and identical except for the square moving independently to the centre. When presented stereoscopically, the square appeared to move towards the observers, as shown in the binocular fused image.

size. The aftereffects produced by adaptation to the visual stimulus alone are smaller than those produced by auditory adaptation (A+, A-), although there is no significant difference between these two combinations. When auditory and visual adapting stimuli were combined in a compatible direction (A+V+, A-V-), the magnitude of the aftereffect increased significantly. This exceeds the sum of the two aftereffects in auditory-alone and visual-alone conditions. Almost the same magnitude was obtained in the auditory-alone condition and the opposite-direction condition (A+V-, A-V+), indicating that, when the directions were opposite, the aftereffect resulted from auditory adaptation, with little or no effect from visual adaptation. These properties—cross-modal enhancement for congruent inputs and no effect for incongruent inputs—are consistent with previously reported results in which aftereffects were not used¹⁶.

In our second experiment, we attempted to confirm that just the visual motion-in-depth cue influences the auditory aftereffect. The left and right eyes each viewed a square moving in opposite directions on the flat screen. In the binocular fused image, the square appeared to move in stereoscopic depth either towards or away from the observers (Fig. 2). A fixed-size square changing in retinal disparity between -20 and +20 min at $\pm 10 \text{ min s}^{-1}$ was used as the visual adapting stimulus. This disparity range corresponds to a distance of about 40 cm in the three-dimensional world. The auditory adapting and test stimuli were identical to those in experiment (1). The changing-loudness aftereffect was measured after adaptation to each of the eight adapting conditions. In these results, we still see the same pattern of the visual effect as in experiment (1) (Fig. 3), although the effects of visual adaptation are relatively small. The auditory aftereffect observed after adaptation to visual stimulus only (V+, V-) was statistically significant, and the magnitude of the aftereffect tended to increase when the directions were the same (A+V+, A-V-); moreover, the magnitude obtained in the opposite-direction condition (A+V-, A-V+) was almost the same as that in auditory-alone condition (A+, A-). The visual effect that is smaller than seen with experiment (1) can be attributed to the weak sensation of motion in depth. The disparity

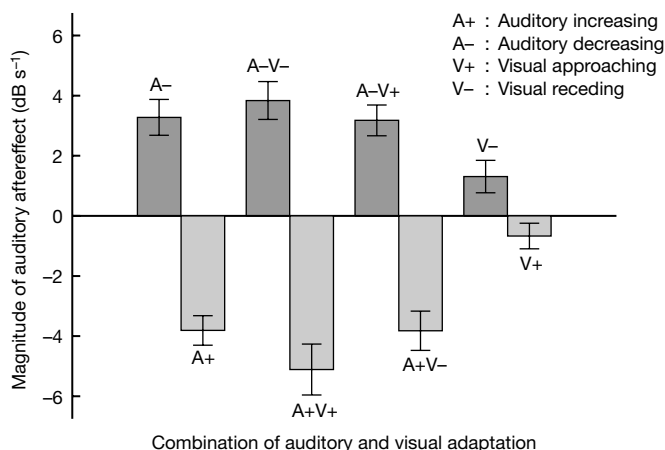


Figure 3 Effect of visual adaptation to changing-disparity on the auditory changing-loudness aftereffect. Averages of the ten subjects are shown. Error bars indicate the standard error of the mean. The aftereffect following adaptation to visual changing disparity alone (V+, V-) was still statistically significant ($P < 0.05$) and the other three combination conditions also produced the significant aftereffect ($P < 0.01$). The absolute value of the magnitude were subjected to two-way ANOVA (adapting direction $\times$ adapting combination). There was a significant main effect of the adapting combination ($F_{3,27} = 17.925$, $P < 0.01$). The aftereffect in the visual alone condition (V+, V-) was weaker than the other combination conditions (Tukey's HSD, $P < 0.01$). The same-direction condition (A+V+, A-V-) increased the aftereffect slightly, although it is not statistically significant.

cue is available only for a relatively short distance and the range of change is limited because the stimulus of a too large disparity is perceived as a double image. Thus, the changing-disparity stimulus should produce a relatively weak impression of motion in depth.

Both visual size- and disparity-changing provide cues for motion in depth and both affected the auditory aftereffect in the same way. Our conclusion is that the visual motion-in-depth cue causes the auditory aftereffect. The auditory system 'senses' sound-source movement in depth from the visual cue of motion in depth, even if there is no sound. It would seem ecologically advantageous for the auditory system to utilize the visual cue. Spatial resolution of the visual system is superior to that of the auditory system. Moreover, it is thought that auditory depth or distance localization, which relies on intensity, reverberation and spectral information¹⁷, is poor compared with horizontal localization in which interaural time and intensity differences can be used. It has been reported that auditory motion perception is more strongly influenced by visual cues in depth than in the horizontal plane⁴.

For spatial events, it is known that visual perception dominates auditory perception¹⁶. In our third experiment, we examined whether the visual changing-size aftereffect¹³ is affected by adaptation to auditory intensity-changing. The changing-size aftereffect is an experience in which, after adaptation to changing size, a subsequently viewed square appears to move in depth in the opposite direction. The test stimulus was a square changed in size from 1 deg $\times$ 1 deg. The auditory and visual adapting stimuli were identical to those in experiment (1). We measured the visual aftereffect after adaptation to each of the eight adapting conditions. There were no aftereffects in auditory-alone conditions (A+, A-) but similar magnitudes of the visual aftereffects were obtained in the other conditions (Fig. 4), suggesting that the auditory adaptation has no effect on the visual changing-size aftereffect. We reconfirm the visual superiority for spatial perception by using the aftereffects.

Most previous auditory-visual interaction studies have examined the visual effect on auditory perception by simultaneous presentation of visual and auditory stimuli. However, by measuring the aftereffect, we found that the presentation of visual stimulus only affects the successive auditory perception. The result suggests that the auditory system, which has been considered to be auditory-

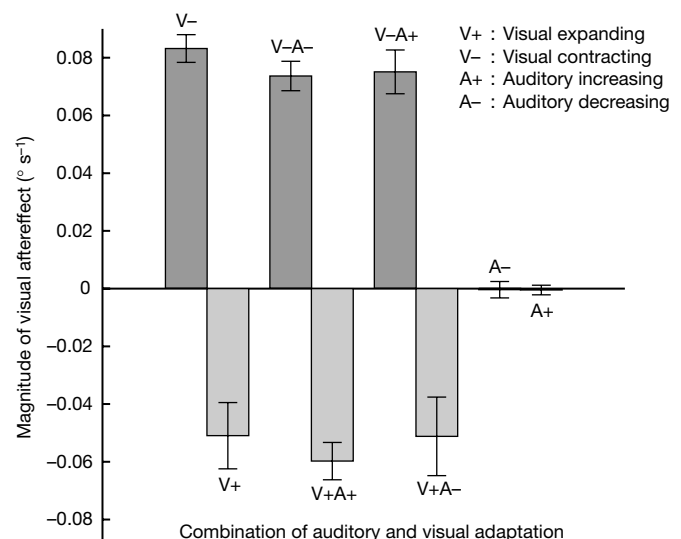


Figure 4 The visual changing-size aftereffect for four combinations in expanding and shrinking directions. Averages of the seven subjects are shown. Error bars indicate the standard error of the mean. Adaptation to auditory changing intensity alone (A+, A-) produced little or no visual aftereffect and the other three combination conditions produced a similar magnitude of the aftereffect.

specific, responds also to visual motion in depth. There are no known direct pathways between auditory and visual cortices, so we suggest that the auditory aftereffect produced by visual adaptation is due to a multimodal process which combines auditory and visual inputs and which subsequently projects back to the auditory motion-in-depth process. Recent evidence from brain-imaging indicates that brain areas which have been considered to be 'unimodal' are activated or modulated by inputs from other sensory modalities^{18–20}. It is suggested that such activation in unimodal areas rely on feedback projections from multimodal areas^{21,22}. In the ventriloquism effect, we actually hear the sound as though it comes from the location of the visual stimulus. To change conscious hearing, a change of neural activities in the auditory area may be necessary.

These results also show properties consistent with previous auditory–visual interaction studies: enhancement for congruent inputs, no effect for incongruent inputs, and visual superiority for spatial perception. The cross-modal aftereffect we have observed may therefore provide a new way to investigate the interaction of the auditory and visual systems for spatial perception and, more generally, interactions between sensory modalities. □

Methods

Ten students participated in experiments (1) and (2). Seven of them also participated in experiment (3). All subjects were naive except for one in experiment (2). All had normal or corrected-to-normal visual acuity, and had no hearing problems. The auditory and visual stimuli were generated by MATLAB with psychophysics toolbox extension^{23,24} on Apple PowerMacintosh. The sampling rate of the sound signal was 44.1 kHz and quantization was 16 bit. The auditory stimuli were converted to analog signals using the computer internal audio interface and presented to both ears through headphones (Sennheiser HDA200). The visual stimuli were presented on a display (SONY GDM-17SE2T, 75-Hz refresh). The experiment was conducted in a darkroom. The subject's head was fixed on a chin rest. The viewing distance was 140 cm. In experiments (1) and (3), viewing was binocular and, in experiment (2), the subjects observed the stimuli through liquid-crystal shutter goggles (Stereographics CrystalEyes).

Procedure

In the adapting conditions, the adapting stimulus was presented 60 times at 200-ms intervals. After the last adapting stimulus offset for 200 ms, a test stimulus was presented, and the subject was instructed to indicate whether this test stimulus appeared to grow louder or lower (experiments (1) and (2)) or grow bigger or smaller (experiment (3)). After the subject's response was given, the adapting stimulus was presented an additional five times to maintain the adaptation, followed by another test stimulus and response interval, and so on. Because a steady test tone after adaptation to a decreasing sound will be perceived as having increasing loudness, a test tone decreasing at a certain rate should be perceived as steady. The double-staircase method²⁵ was used to measure such points of subjective steadiness of the test stimulus, that is, the velocity of the physical change in level (experiments (1) and (2)) or in size (experiment (3)) necessary for the test stimulus to be perceived as steady. After the subjects' response, the velocity of the test stimulus was varied by 1 dB s⁻¹ step in experiments (1) and (2) and by 1 min s⁻¹ step in experiment (3). The point of subjective steadiness was calculated by averaging the last five (out of ten) reversal points of each staircase series. Before each adapting session, the point of subjective steadiness was measured without adaptation. The difference between the point of subjective steadiness in adaptation and no-adaptation conditions (no-adaptation minus adaptation) was taken as the magnitude of the aftereffect.

Received 8 November; accepted 11 December 2001.

1. Jack, C. E. & Thurlow, W. R. Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Percept. Motor Skill* **37**, 967–979 (1973).
2. Bertelson, P. & Radeau, M. Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept. Psychophys.* **29**, 578–584 (1981).
3. Mateeff, S., Hohnsbein, J. & Noack, T. Dynamic visual capture: apparent auditory motion induced by a moving visual target. *Perception* **14**, 721–727 (1985).
4. Kitajima, N. & Yamashita, Y. Dynamic capture of sound motion by light stimuli moving in three-dimensional space. *Percept. Motor Skill* **89**, 1139–1158 (1999).
5. Wade, N. J. A selective history of the study of visual motion aftereffects. *Perception* **23**, 1111–1134 (1994).
6. Mather, G., Verstraten, F. & Anstis, S. *The Motion Aftereffect: A Modern Perspective* (MIT Press, Cambridge, Massachusetts, 1998).
7. Grantham, D. W. Motion aftereffects with horizontally moving sound sources in the free field. *Percept. Psychophys.* **45**, 129–136 (1989).
8. Dong, C. J., Swindale, N. V., Zakarauskas, P., Hayward, V. & Cynader, M. S. The auditory motion aftereffect: its tuning and specificity in the spatial and frequency domains. *Percept. Psychophys.* **62**, 1099–1111 (2000).
9. Shu, Z. J., Swindale, N. V. & Cynader, M. S. Spectral motion produces an auditory after-effect. *Nature* **364**, 721–723 (1993).
10. Reinhardt-Rutland, A. H. & Anstis, S. Auditory adaptation to gradual rise or fall in intensity of a tone. *Percept. Psychophys.* **31**, 63–67 (1982).

11. Reinhardt-Rutland, A. H. Increasing-loudness aftereffect following decreasing-intensity adaptation: spectral dependence in interotic and monotonic testing. *Perception* **27**, 473–482 (1998).
12. Kayahara, T. Aftereffect of adaptation to uni-direction frequency change: evidence for selective processing mechanism. *Acoust. Sci. Technol.* **22**, 49–51 (2001).
13. Regan, D. & Beverley, K. I. Illusory motion in depth: aftereffect of adaptation to changing size. *Vision Res.* **18**, 209–212 (1978).
14. Cumming, B. in *Visual Detection of Motion* (eds Smith, A. T. & Snowden, R. J.) 333–366 (Academic, London, 1994).
15. Ehrenstein, W. H. & Reinhardt-Rutland, A. H. A cross-modal aftereffect: auditory displacement following adaptation to visual motion. *Percept. Motor Skill* **82**, 23–26 (1996).
16. Welch, R. B. & Warren, D. H. in *Handbook of Perception and Human Performance*. Vol. 1 Sensory Processes and Perception (eds Boff, K. R., Kaufman, L. & Thomas, J. P.) Ch. 25 (Wiley, New York, 1986).
17. Blauert, J. *Spatial Hearing: the Psychophysics of Human Sound Localization* Revised edn (MIT Press, Cambridge, Massachusetts, 1997).
18. Calvert, G. A. et al. Activation of auditory cortex during silent lipreading. *Science* **276**, 593–596 (1997).
19. Macaluso, E., Frith, C. D. & Driver, J. Modulation of human visual cortex by crossmodal spatial attention. *Science* **289**, 1206–1208 (2000).
20. Lewis, J. W., Beauchamp, M. S. & DeYoe, E. A. A comparison of visual and auditory motion processing in human cerebral cortex. *Cereb. Cortex* **10**, 873–888 (2000).
21. Driver, J. & Spence, C. Multisensory perception: beyond modularity and convergence. *Curr. Biol.* **10**, R731–R735 (2000).
22. Calvert, G. A., Campbell, R. & Brammer, M. J. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.* **10**, 649–657 (2000).
23. Brainard, D. H. The psychophysics toolbox. *Spatial Vision* **10**, 433–436 (1997).
24. Pelli, D. G. The Video Toolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision* **10**, 437–442 (1997).
25. Cornsweet, T. N. The staircase-method in psychophysics. *Am. J. Psychol.* **75**, 485–491 (1962).

Acknowledgements

We thank D. Levi, H. Akutsu and D. Erickson for helpful comments and suggestions.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.K. (e-mail: kitagawa-norimichi@c.metro-u.ac.jp).

Embryonic assembly of a central pattern generator without sensory input

Maximiliano L. Suster* & Michael Bate

Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK

Locomotion depends on the integration of sensory information with the activity of central circuitry, which generates patterned discharges in motor nerves to appropriate muscles^{1,2}. Isolated central networks generate fictive locomotor rhythms (recorded in the absence of movement), indicating that the fundamental pattern of motor output depends on the intrinsic connectivity and electrical properties of these central circuits^{3,4}. Sensory inputs are required to modify the pattern of motor activity in response to the actual circumstances of real movement. A central issue for our understanding of how locomotor circuits are specified and assembled is the extent to which sensory inputs are required as such systems develop⁵. Here we describe the effects of eliminating sensory function and structure on the development of the peristaltic motor pattern of *Drosophila* embryos and larvae. We infer that the circuitry for peristaltic

* Present address: Department of Zoology, University of Toronto at Mississauga, Mississauga, Canada L5L 1C6.

crawling develops in the complete absence of sensory input; however, the integration of this circuitry into actual patterns of locomotion requires additional information from the sensory system. In the absence of sensory inputs, the polarity of movement is deranged, and backward peristaltic waves predominate at the expense of forward peristalsis.

The *Drosophila* larva moves over substrates by peristaltic crawling⁶. In forward locomotion, coordinated waves of muscle contraction pass along the body segments from posterior to anterior. In backward movement the contraction wave is reversed and passes from anterior to posterior. Larvae tend to move forward, but briefly move backwards in response to a mechanical stimulus to the anterior segments⁷. The peristaltic movements of the larva are preceded developmentally by vigorous movements of the embryo during late stages of embryogenesis⁸. These embryonic movements consist of repeated forward and backward waves of peristalsis within the egg shell, which culminate in a sequence of head swings, the puncturing of the egg shell, and the hatching of the larva by a series of forward peristaltic waves that release it from the enclosing vitelline membrane⁹. In the absence of neurotransmitter release, or if the central nervous system (CNS) is selectively removed with a toxin, peristalsis fails ($n = 0/10$ embryos each, see Methods). Thus embryonic and larval peristalsis, like other repetitive movements, depends on the coordinated output of central pattern generating (CPG) circuits, and these circuits mature and start to function during the second half of embryogenesis.

In *Drosophila*, the larval sensory system (peripheral nervous system) consists of a stereotyped array of external sensory organs, chordotonal organs, and multidendritic neurons. To assess whether sensory input is required for the normal development of central locomotor circuits, we used the Gal4/upstream activating sequence (UAS) system¹⁰, and a specific Gal4 driver¹¹ (Fig. 1) to target the expression of tetanus toxin light chain (TeTxLC) to the neurons of the peripheral nervous system during embryogenesis, thus blocking neurotransmitter release from these cells¹². In this experiment, a transgene coding for the yeast transcriptional activator Gal4 is driven by a peripheral nervous system-specific enhancer to induce expression of a TeTxLC transgene¹⁰. We then quantified embryonic movements from 16 h after egg laying until hatching (21 h) using

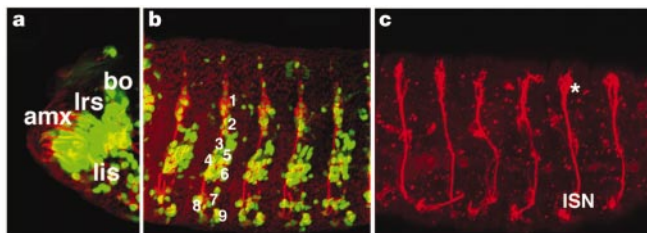


Figure 1 Genetic elimination of sensory function or structure in the *Drosophila* embryonic peripheral nervous system. **a**, **b**, P0163-Gal4; UAS-*GFPN-lacZ* embryos were double stained with anti- β -gal and 22C10 antibodies¹¹ to check that all neurons of the peripheral nervous system (revealed by cytoplasmic marker 22C10, in red) expressed Gal4 (revealed by nuclear β -gal, in green). **a**, All sensory organs in the head such as Bolwig's organ (bo), labral sensory complex (lrs), labial sensory complex (lis), and antennomaxillary complex (amx) showed Gal4 expression. **b**, All sensory neurons (groups numbered 1–9)¹³ showed expression of Gal4. Expression is completely restricted to the peripheral nervous system, except for the ring gland and some scattered glia of the CNS (not shown). Expression of TeTxLC in the embryonic ring gland (using OK236 and P0206 Gal4 lines) or CNS glia (OK75, 8, 72, 83, 259 and 326 Gal4 lines) did not affect the embryonic or larval motor pattern. **c**, Sensory neurons are missing in *sens^{D(GAL)1228/4}* embryos. Homozygous mutant embryos were stained with 22C10 antibody to reveal the peripheral nervous system. All sensory neurons are missing except for multidendritic neurons, which were affected variably (marked by an asterisk). ISN, inter-segmental nerve. Magnification, about $\times 120$.

time-lapse videomicroscopy to monitor the positions of the segmental denticle bands during peristaltic waves. The denticle bands serve as frictional anchoring points during crawling, and their movements provide a measure of the underlying phases of muscle contraction and relaxation during peristalsis. Control and experimental embryos generate well coordinated waves of forward and backward peristalsis (Fig. 2 and Supplementary Information), although the overall period for a forward wave is longer, and there is a marked reduction in the frequency of forward peristalsis in embryos developing without input from the peripheral nervous system (Fig. 3a, b). We conclude that the central circuits required for the coordinated movements of peristalsis differentiate successfully in the absence of sensory inputs, but that the overall polarity of movement is abnormal. Although such embryos hatch successfully from the egg case, the behaviour of the hatched larvae is very abnormal: they have a low frequency of forward movements and make repeated backward movements and frequent head swings (Fig. 4). As expected, these larvae are completely insensitive to mechanical stimuli ($n = 0/200$ responded to touch⁷).

Sensory axons enter the CNS at an early stage in its development¹³. Thus, although our experiments indicate that input from

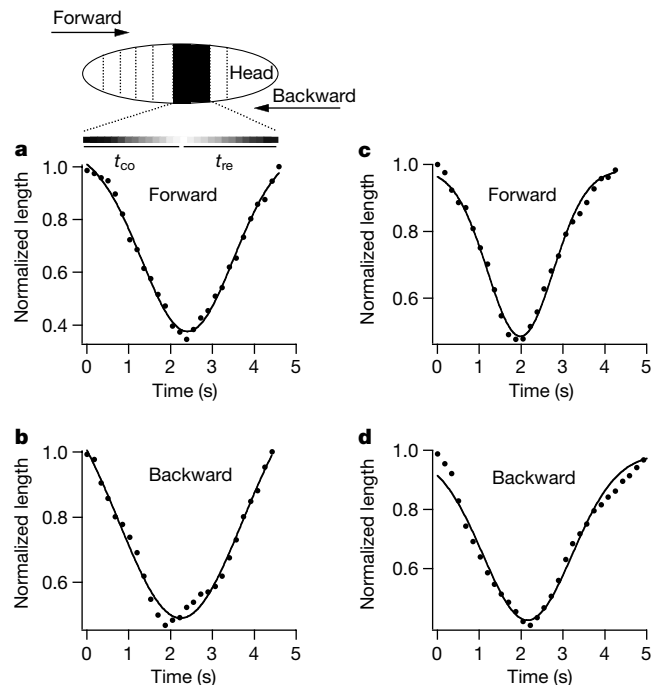


Figure 2 Peristalsis in embryos expressing active and inactive TeTxLC in the peripheral nervous system. The combined length of segments A2 and A3 was measured for each embryo (shaded segments), normalized (divided by the maximum length) and plotted against time (s) for individual waves of peristalsis. Time of contraction (t_{co}) and time of relaxation (t_{re}) were defined as the time between maximum and minimum length (contraction), and vice versa (relaxation). **a**, **b**, Embryos expressing inactive TeTxLC in the peripheral nervous system show forward (**a**) and backward (**b**) waves with contraction and relaxation phases of approximately equal duration ($t_{co} = 2.55 \pm 0.24$ s; $t_{re} = 2.48 \pm 0.33$, $n = 10$, for forward waves). **c**, **d**, Contraction and relaxation phases in embryos expressing active TeTxLC during forward (**c**) and backward (**d**) peristalsis, do not differ from embryos expressing inactive TeTxLC ($t_{co} = 2.50 \pm 0.25$ s; $t_{re} = 2.59 \pm 0.41$ s, $n = 7$, for forward waves). In general, however, the completion of forward waves took longer in embryos expressing active ($n = 15$) than inactive ($n = 33$) TeTxLC (14.4 ± 1.6 s compared with 9.7 ± 0.7 s, respectively, $P < 0.01$). There was no difference in the mean duration of backward waves (11.9 ± 1 s, inactive TeTxLC compared to 12.3 ± 0.7 s, active TeTxLC, $n > 18$ each, ANOVA). Sample movies of peristalsis from active and inactive embryos expressing TeTxLC are provided in the Supplementary Information.

sense organs is not required for the successful assembly of the central pattern generator, the sensory axons and their terminals might be cellular components that are essential for the proper development of locomotor circuits. They might be required either as a substrate for the growth of other axons and dendrites, or as a source of signals for the proliferation¹⁴ or differentiation¹⁵ of other neurons. To resolve this issue we repeated our experiments using a null mutation in the *senseless* gene, *sens*^{Df(3L)1228/4}. In the absence of the Senseless protein, sensory neurons, with the exception of some multidendritic neurons, fail to develop (39.1% of the four most dorsal multidendritic neurons are present, $N = 5$ embryos, 35 hemisegments)¹⁶. Thus in *sens*^{Df(3L)1228/4} embryos the axons and terminals of most sensory neurons (Fig. 1c) are lacking from the CNS throughout its development, and we can use this aspect of the *sens*^{Df(3L)1228/4} phenotype to ask whether there is any requirement for sensory neurons and their axons in the normal development of locomotor circuitry. The consequences of the *sens*^{Df(3L)1228/4} mutation for embryonic movement closely resemble the effects of expressing tetanus throughout the sensory system: the embryos

generate coordinated peristaltic waves, but the frequency of forward waves is markedly curtailed (Fig. 3a, b). Thus the axons and terminals of most sensory neurons are not required for the normal development of locomotor circuitry. However, these results reinforce the conclusion that sensory input is required to establish the polarity of movement both in the embryo and the larva.

The finding that the occurrence of forward waves is preferentially sensitive to the loss of sensory input may indicate that episodes of forward and backward peristalsis are controlled in different ways. To look at this issue in greater detail, we examined the temporal distribution of forward and backward peristaltic waves during the period of late embryonic movement. In wild-type embryos, the distribution of forward and backward waves is clearly different (Fig. 5a, b). Most forward waves occur in bursts (waves clustered in time intervals less than 74 s), whereas the remainder (forward and backward) occur sporadically. Although the temporal distribution of backward waves is unaffected by the removal of sensory function (Fig. 5b, d), bursts of forward waves are virtually eliminated (compare Fig. 5c with a). The fact that sporadic waves of forward motion still occur in embryos devoid of sensory input shows that the machinery for forward peristalsis has been assembled correctly, but suggests that there may be a special requirement for the peripheral nervous system to initiate bursts of forward movement.

Detailed studies of other locomotor systems show that their performance depends not only on the proper differentiation of CPG neurons, but also on descending inputs that may be modulatory in their effects or lead to the initiation or blocking of motor activity^{17,18}. Thus, in *Xenopus*, inputs from descending interneurons are essential for the integration of the swimming CPG

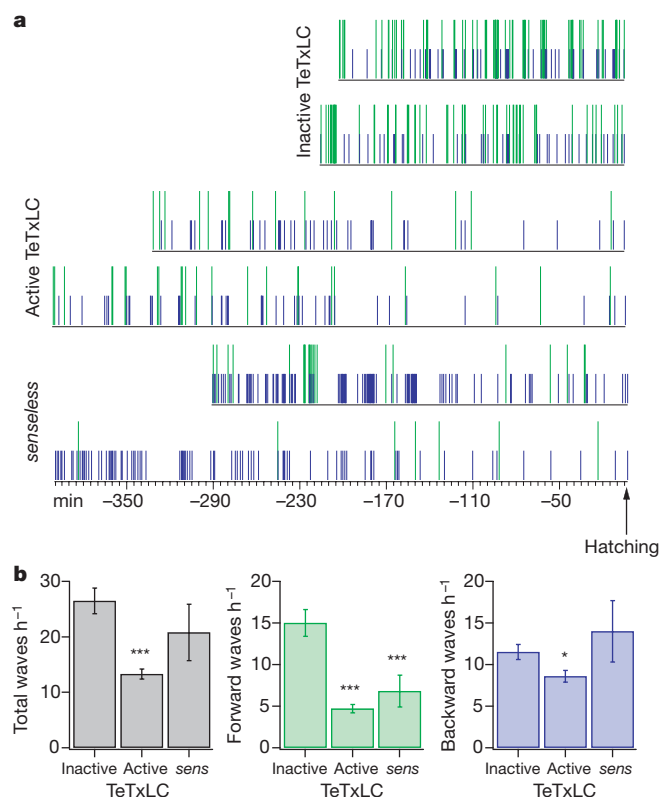


Figure 3 Elimination of sensory input disrupts episodes of forward but not backward peristaltic waves. **a**, Peristaltic motor pattern in the *Drosophila* embryo. Full waves of forward (green lines) and backward (blue lines) peristalsis were recorded from their first occurrence until hatching (see Methods). In general, this interval is longer in embryos lacking sensory input. Control P0163-Gal4; TNT-VIF (inactive TeTxLC) embryos showed a robust pattern of waves that was slightly biased towards forward movement. In contrast, P0163; TNT-G (active TeTxLC) embryos showed a remarkable loss of forward waves, whereas backward waves seemed unaffected. *sens*^{Df(3L)1228/4} embryos (*senseless*) show few, if any, forward waves of peristalsis, whereas in some, bursts of backward peristalsis were more frequent than in controls. **b**, Total forward and backward waves were counted and compared to the inactive TeTxLC control. Total waves per hour were diminished only in embryos expressing active TeTxLC (triple asterisk, $P < 0.0001$). Forward waves per hour were significantly reduced in embryos expressing active TeTxLC and embryos carrying the *sens*^{Df(3L)1228/4} mutation (triple asterisk, $P < 0.0001$). Backward waves were only slightly reduced in embryos expressing active TeTxLC (asterisk, $P < 0.05$). $n \approx 8$ per bar.

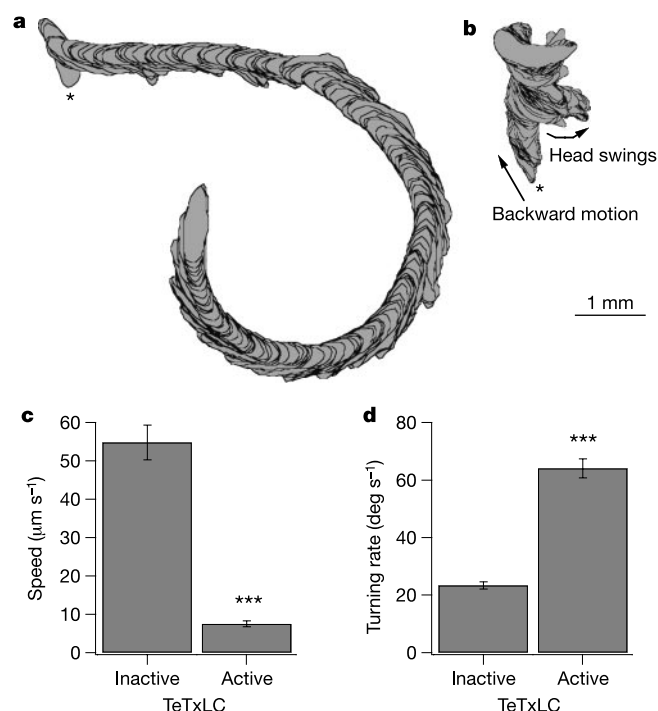


Figure 4 Expression of tetanus toxin (TeTxLC) in the embryonic peripheral nervous system results in severely disrupted larval locomotion. **a**, **b**, Crawling episodes (3 min) of first instar larvae (0–1 h old) expressing inactive or active TeTxLC are shown as perimeter stacks overlaid for the entire crawling episode. The asterisk indicates the first frame of each movie. Larvae expressing active TeTxLC (**b**) show a high frequency of backward movements and repetitive head swings compared with the inactive TeTxLC control (**a**). **c**, **d**, Parameters of crawling in larvae expressing TeTxLC. Larvae expressing active TeTxLC show a marked reduction (triple asterisk, $P < 0.0001$) in average speed (**c**) and increase in mean turning rate (**d**). $n = 10$ for inactive TeTxLC; $n = 17$ for active TeTxLC.

into two behaviour patterns, which are acquired by the embryo as it matures and hatches: activation of swimming in response to lights off, mediated by the dorsal pineal eye, and inhibition of swimming in response to pressure on the anterior cement gland¹⁹. Our results with *Drosophila* indicate that sensory inputs may also be required to integrate the activity of the peristaltic CPG into actual patterns of larval locomotion; however, we cannot exclude the possibility that in normal larvae, sensory inputs simply act to enhance the excitability of the CPG.

Here we have shown that sensory inputs are not essential for the differentiation and maturation of motor circuitry in *Drosophila*. Tetanus expression removes the input of all sensory neurons (although electrical synapses may remain intact), and *sens^{Df(3L)1228/4}* removes the input of most sensory neurons. In addition, the *senseless* mutation removes a substantial cellular component from the developing CNS, namely the axons and terminals of sensory neurons. Yet, in both cases, embryos succeed in assembling circuitry necessary for the generation of well coordinated waves of forward and backward peristalsis. This is consistent with previous experiments in which immobilization of *Xenopus* embryos with anaesthetics resulted in no obvious disruptions of motor development²⁰, and points to a high degree of autonomy in the differentiation and maturation of neurons contributing to CPG circuits underlying larval locomotion in *Drosophila*.

The fact that sensory input does not seem to be required does not mean that activity has no part in the development of the locomotor system. In general, embryonic motor systems generate precocious activity that is manifested as patterns of embryonic movement^{21–23}.

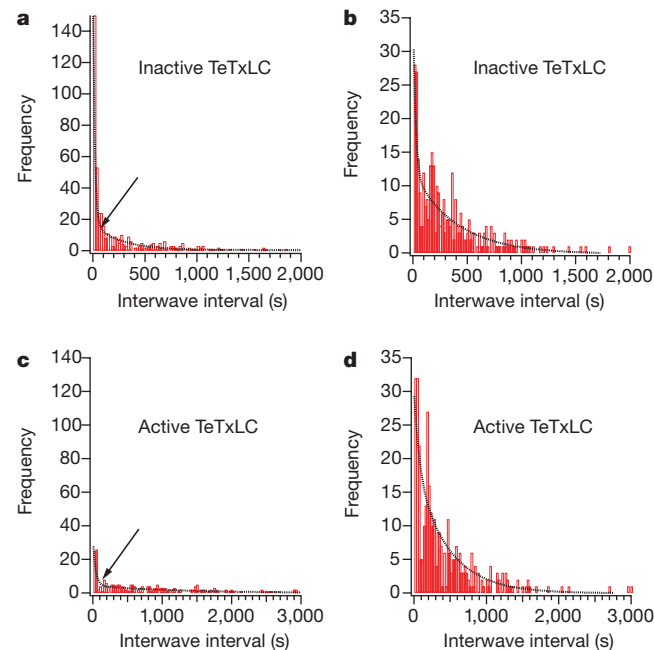


Figure 5 Bursts of forward peristalsis but not sporadic forward waves are eliminated by removal of sensory input. The time interval between consecutive forward waves (interwave interval in seconds) or backward waves was computed and plotted against frequency for embryos expressing inactive TeTxLC (**a, b**) and active TeTxLC (**c, d**) in the peripheral nervous system. **a**, Forward waves show a double exponential distribution (made up of two exponential curves) with most waves occurring in bursts, whereas the remainder occur at long intervals. The arrow points to the meeting of the two exponential curves (see Methods), which can be used to distinguish waves that occur in bursts (interwave interval less than 74 s) from those that occur sporadically (interval greater than 74 s). **b**, Backward waves show a greater spread in their distribution with fewer waves occurring in bursts than forward waves. **c**, In the absence of sensory input, sporadic forward waves are still present but those which occur in bursts are missing. **d**, The distribution of backward waves is unchanged in the absence of sensory input.

The developmental significance of this motor activity is not clear, but one possibility is that, as in sensory systems²⁴, activity in the motor system is a prerequisite for the proper maturation and refinement of connections. If the characteristics of the CPG seem, as in this case, to be determined autonomously by the differentiation of central circuit components alone, it may well be that the acquisition of appropriate connectivity and electrical characteristics is, in part, an activity-dependent process. In this case the properties of neurons would be dependent on and adjusted in response to actual patterns of input from their central synaptic partners. Such adjustments have been documented in other systems²⁵, and there is already evidence from *Drosophila* that larval motor neurons fail to develop their normal membrane characteristics in the absence of synaptic input²⁶.

Although the fundamental circuitry required to generate a coordinated forward or backward sequence of contractions is assembled successfully in embryos lacking sensory inputs, the overall behaviour of these animals is highly abnormal. An interesting question that can now be addressed is whether this abnormality reveals an acute requirement for sensory inputs in organizing movement episodes, or whether there is a more fundamental requirement for the sensory system to be active as this organizing machinery develops. More generally, the discovery that neurons of the CNS differentiate autonomously to produce the CPG for larval peristalsis in *Drosophila* provides us with an opportunity to investigate how such embryonic motor systems are specified and assembled. □

Methods

Fly strains

Flies were kept on standard media at $25 \pm 1^\circ\text{C}$. The *sens^{Df(3L)1228/4}* mutation was provided by H. Bellen¹⁶ and balanced over *TM3, Ser^{GFP}* to identify homozygous mutant animals. P0163-Gal4 has been described¹¹, and was obtained from the Umea Stock Centre. We obtained second-chromosome viable TeTxLC stocks, UAS-TNT-VIF (inactive) and UAS-TNT-G (active) from S. Sweeney¹². For pan-neuronal expression of TeTxLC, female C155-Gal4 (ref. 27) and *elav*-Gal4 (ref. 28) were crossed independently to UAS-TeTxLC males. For ablation of the nervous system, C155 was crossed to UAS-*ricin/CyO* (ref. 29) males.

Computer-assisted analysis of crawling behaviour

First instar 0–1-h-old larvae were filmed freely moving on an agar substrate for 3 min using a CCD (charge-coupled device) monochrome Hitachi KP-ME1 camera. Movies of crawling were captured at 1 frame per second on a Macintosh G3 computer with a Scion LG3 framegrabber (Scion Corp.). Movies were analysed using the Dynamic Image Analysis System (Solltech). We obtained speed and turning rate automatically from DIAS³⁰.

Video recording of embryonic peristalsis

Embryos were collected overnight on agar plates supplemented with apple juice and dechlorinated (1:1 bleach:water), washed and selected for video recording. Eight to ten embryos were placed in a petri dish under $1 \times \text{PBS}$ and video recorded on top of an evenly fluorescent light box (Kenro) using a Wild stereo microscope. Videos of movement were recorded from the initiation of peristalsis (at the stage at which embryos fill their tracheae) to hatching, using a time-lapse VHS video recorder (Panasonic AG-6040) at $21–25^\circ\text{C}$. Forward and backward waves of peristalsis were reviewed at $8\times$ real speed, and scored on a monitor. The time of wave initiation (± 4 s) and type of wave (backward or forward) were recorded on a spreadsheet using Excel (Microsoft). *sens^{Df(3L)1228/4}* embryos do not hatch because they fail to break the vitelline membranes with their egg teeth. Therefore, peristaltic movements in these embryos were recorded for 6–8 h after the beginning of coordinated peristalsis. To measure kinematic parameters of peristalsis (for example, wave duration), embryos were video recorded individually under a $\times 10$ objective on a Zeiss Axioskop microscope using a CoolFire digital monochrome CCD camera (Empix Imaging). We captured 300 frames (672×296 pixels) at approximately 6 frames per second on a PC computer using the Northern Elite image analysis system (Empix Imaging).

Statistics and graphics

We documented the pattern of peristaltic waves in embryos in two ways. First, colour-coded lines representing waves of forward (green) or backward (blue) peristalsis were plotted on the same time axis (min), and aligned at hatching. Second, total number of waves, forward waves and backward waves per hour were computed. Numbers of waves per hour were compared using a Mann–Whitney *U*-test. Double exponential curves were fitted to the distributions of time intervals (s) between forward or backward peristaltic waves using Igor Pro 3.16 (Wavemetrics). Larval speed and turning rate measurements were compared using one-way analysis of variance (ANOVA) in SPSS 10 (SPSS). We obtained kinematic parameters of peristalsis manually using QuickTime 4.1 and NIH Image 1.61.

Received 1 August 2001; accepted 11 January 2002.

- Dickinson, M. H. *et al.* How animals move: an integrative view. *Science* **288**, 100–106 (2000).
- Pearson, K. G. Proprioceptive regulation of locomotion. *Curr. Opin. Neurobiol.* **5**, 786–791 (1995).
- Marder, E. & Calabrese, R. L. Principles of rhythmic motor pattern generation. *Physiol. Rev.* **76**, 687–717 (1996).
- Grillner, S. Neurobiological bases of rhythmic motor acts in vertebrates. *Science* **228**, 143–149 (1985).
- Sanes, D. H., Reh, T. A. & Harris, W. A. *Development of the Nervous System* (Academic, San Diego, 2000).
- Berrigan, D. & Pepin, D. J. How maggots move: allometry and kinematics of crawling in larval diptera. *J. Insect Physiol.* **41**, 329–337 (1995).
- Kernan, M., Cowan, D. & Zucker, C. Genetic dissection of mechanosensory transduction: mechanoreception-defective mutations of *Drosophila*. *Neuron* **12**, 1195–1206 (1994).
- Kalish, N. The effect on development of a lethal deficiency in *Drosophila melanogaster*: with a description of the normal embryo at the time of hatching. *Genetics* **24**, 244–270 (1939).
- Siekhous, D. E. & Fuller, R. S. A role for *amontillado*, the *Drosophila* homolog of the neuropeptide precursor processing protease PC2, in triggering hatching behaviour. *J. Neurosci.* **10**, 6942–6954 (1999).
- Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
- Hummel, T., Krukkert, K., Roos, J., Davis, G. & Klämbt, C. *Drosophila* Futsch/22C10 is a MAP1B-like protein required for dendritic and axonal development. *Neuron* **26**, 357–370 (2000).
- Sweeney, S. T., Broadie, K., Keane, J., Niemann, H. & O’Kane, C. J. Targeted expression of tetanus toxin light chain in *Drosophila* specifically eliminates synaptic transmission and causes behavioural deficits. *Neuron* **14**, 341–351 (1995).
- Jan, Y. N. & Jan, L. Y. in *The Development of Drosophila melanogaster* (eds Bate, M. & Martinez-Arias, A.) 1207–1244 (Cold Spring Harbor Laboratory Press, New York, 1993).
- Huang, Z. & Kunes, S. Hedgehog, transmitted along retinal axons, triggers neurogenesis in the developing visual centers of the *Drosophila* brain. *Cell* **86**, 411–422 (1996).
- Schneiderman, A. M., Hildebrand, J. G., Brennan, M. M. & Tumlinson, J. H. Trans-sexually grafted antennae alter pheromone-directed behaviour in a moth. *Nature* **323**, 801–803 (1986).
- Nolo, R., Abbot, L. A. & Bellen, H. J. Senseless, a Zn finger transcription factor, is necessary and sufficient for sensory organ development in *Drosophila*. *Cell* **102**, 349–362 (2000).
- Marder, E. Motor pattern generation. *Curr. Opin. Neurobiol.* **10**, 691–698 (2000).
- Richards, K. S. & Marder, E. The actions of crustacean cardioactive peptide on adult and developing stomatogastric ganglion motor patterns. *J. Neurobiol.* **44**, 31–44 (2000).
- Boothby, K. M. & Roberts, A. The stopping response of *Xenopus laevis* embryos: behaviour, development and physiology. *J. Comp. Physiol.* **170**, 171–180 (1992).
- Haverkamp, L. J. & Oppenheim, R. W. Behavioural development in the absence of neural activity: effects of chronic immobilization on amphibian embryos. *J. Neurosci.* **6**, 1332–1337 (1986).
- Hamburger, V. & Oppenheim, R. Prehatching motility and hatching behavior in the chick. *J. Exp. Zool.* **166**, 171–203 (1967).
- van Mier, P., Armstrong, J. & Roberts, A. Development of early swimming in *Xenopus laevis* embryos: myotomal musculature, its innervation and activation. *Neuroscience* **32**, 113–126 (1989).
- Reynolds, S. A., French, K. A., Baader, A. & Kristan, W. B. Jr Development of spontaneous and evoked behaviors in the medicinal leech. *J. Comp. Neurol.* **402**, 168–180 (1998).
- Katz, L. C. & Shatz, C. J. Synaptic activity and the construction of cortical circuits. *Science* **274**, 1133–1138 (1996).
- Turrigiano, G. G. Homeostatic plasticity in neuronal networks: the more things change, the more they stay the same. *Trends Neurosci.* **22**, 221–227 (1999).
- Baines, R. A., Uhler, J. P., Thompson, A., Sweeney, S. T. & Bate, M. Altered electrical properties in *Drosophila* neurons developing without synaptic transmission. *J. Neurosci.* **21**, 1523–1531 (2001).
- Lin, D. M. & Goodman, C. S. Ectopic and increased expression of Fasciclin II alters motoneuron growth cone guidance. *Neuron* **13**, 507–523 (1994).
- Luo, L., Liao, Y. J., Jan, L. Y. & Jan, Y. N. Distinct morphogenetic functions of similar small GTPases: *Drosophila* Drac1 is involved in axonal outgrowth and myoblast fusion. *Genes Dev.* **8**, 1787–1802 (1994).
- Hildago, A., Urban, J. & Brand, A. H. Targeted ablation of glia disrupts axon tract formation in the *Drosophila* CNS. *Development* **121**, 3703–3712 (1995).
- Wang, J. W. *et al.* Morphometric description of the wandering behaviour in *Drosophila* larvae: aberrant locomotion in Na and K channel mutants revealed by computer-assisted motion analysis. *J. Neurogenet.* **11**, 231–254 (1997).

Supplementary Information accompanies the paper on Nature’s website (<http://www.nature.com>).

Acknowledgements

We thank our colleagues in Cambridge and Toronto for comments on this work, and M. Sokolowski for providing support during its completion. M.L.S. was funded by a Luis Velez Scholarship from the Venezuelan National Academy of Sciences at Clare College, Cambridge. M.B. is a Royal Society Research Professor. This work was supported by a grant from the Wellcome Trust.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.L.S. (e-mail: msuster@credit.erin.utoronto.ca).

***Drosophila* Crumbs is a positional cue in photoreceptor adherens junctions and rhabdomeres**

Shayan Izaddoost*, Sang-Chul Nam*, Manzoor A. Bhat†, Hugo J. Bellen*‡§ & Kwang-Wook Choi*‡||

* Department of Molecular and Cellular Biology; ‡ Program in Developmental Biology; § Department of Molecular and Human Genetics, Howard Hughes Medical Institute; and || Department of Ophthalmology; Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA
† Cardiovascular Research Institute, Department of Medicine, Department of Molecular Cell and Developmental Biology, Mount Sinai School of Medicine, New York, New York 10029, USA

Drosophila Crumbs (Crb) is required for apical–basal polarity and is an apical determinant in embryonic epithelia^{1,2}. Here, we describe properties of Crb that control the position and integrity of the photoreceptor adherens junction and photosensitive organ, or rhabdomere³. In contrast to normal photoreceptor adherens junctions and rhabdomeres, which span the depth of the retina, adherens junctions and rhabdomeres of Crb-deficient photoreceptors initially accumulate at the top of the retina and fail to maintain their integrity as they stretch to the retinal floor. We show that Crb controls localization of the adherens junction through its intracellular domain containing a putative binding site for a protein 4.1 superfamily protein (FERM)^{4,5}. Although loss of Crb or overexpression of the FERM binding domain causes mislocalization of adherens junctions, they do not result in a significant loss of photoreceptor polarity. Mutations in *CRB1*, a human homologue of *crb*, are associated with photoreceptor degeneration in retinitis pigmentosa 12 (RP12) and Leber congenital amaurosis (LCA)^{6–10}. The intracellular domain of *CRB1* behaves similarly to its *Drosophila* counterpart when overexpressed in the fly eye. Our studies may provide clues for mechanisms of photoreceptor degeneration in RP12 and LCA.

The phototransduction machinery of both vertebrate and *Drosophila* photoreceptors is housed in a morphologically unique and conserved apical domain specialization called the rod/cone outer segment and rhabdomere, respectively. Both of these structures arise from the explosive growth of the photoreceptor apical domain, giving rise to stacks of membrane packed with the photopigment rhodopsin^{11–14}. The role of apical proteins in rhabdomere formation is poorly understood.

To address this matter, we studied Crb function in the eye of *Drosophila melanogaster*. The transmembrane protein Crb and its interacting partner, Discs lost (Dlt)¹⁵, co-localize to the apical domain of undifferentiated cells and differentiating photoreceptors throughout eye development (Fig. 1a–c). During the pupal stage, Crb, Dlt and adherens junctions (AJs) involute with the photoreceptor apical domain. AJs anchor photoreceptors to the surface and floor of the retina. Thus, the axes of the photoreceptor apical domains and AJs rotate 90° and are perpendicular to the surface of the retina (Fig. 1a–c).

In the third-instar and pupal stages before rhabdomere formation, Dlt and Crb are juxtaposed to AJs marked with anti-Arm (Fig. 1d, e). As the actin-rich rhabdomeres form, Crb and Dlt localization is confined to the region referred to as the rhabdomere stalk between the rhabdomere and AJ (Fig. 1f–h). To examine the function of Crb in eye morphogenesis, we generated eye-specific *crb*[−] mosaic clones using the *ey-Flp/FRT* system^{16,17} and two separate alleles of *crb*, *crb*^{D88-3} (a hypomorph) and *crb*^{11A22} (null)¹. Neither type of clone showed any detectable level of Crb protein, although *crb*^{11A22} clones show a slightly more severe

phenotype. Longitudinal sections of the *crb*⁻ clones show lack of rhabdomere elongation (compare arrows to arrowheads, Fig. 2a). The rhabdomeres remain at the top of the retina, and appear thicker than wild-type rhabdomeres. Thus, loss of Crb results in failure of rhabdomere elongation.

To determine the initial defect causing short rhabdomeres, *crb*⁻ mid-pupal-stage eyes were dissected and stained for Crb and Arm. A longitudinal three-dimensional reconstruction of photoreceptor AJs in a photoreceptor cluster is shown in Fig. 2b and c, with the apical domain of the photoreceptors pointing towards the centre of the cluster (modelled in Fig. 1a–c). Wild-type clusters contain smooth and well-defined AJs spanning the length of the retina from the surface to the floor (Fig. 2b). The *crb*⁻ cluster shows an accumulation of Arm staining at the top of the cluster (brackets,

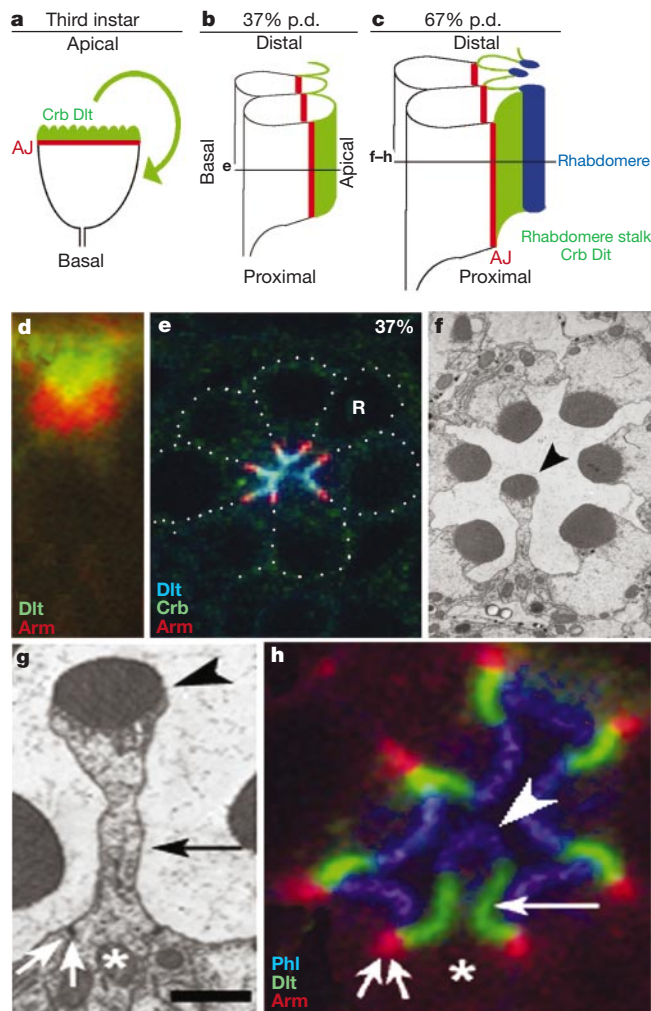


Figure 1 Dlt and Crb are localized to the rhabdomere stalk. The colour of each protein is marked at the bottom of all fluorescence images. **a**, Third-instar eye disc. Apical domains of photoreceptors (green) face the retinal surface and are held together by the AJ (red). **b**, 37% pupal development (p.d.). Apical domains involute and face each other, without breaking their AJ. **c**, 67% p.d. The distal apical domain is anchored by AJs to the retinal surface while the proximal apical domain is bound by AJs to the retinal floor. **d**, Longitudinal view of a third-instar eye disc stained for Dlt and Arm. **e**, Tangential section of a photoreceptor cluster at 37% p.d. stained for Dlt, Crb and Arm. The position of the section is shown in **b**. White dots, photoreceptor (R) basolateral membrane. **f**, **g**, Transmission electron micrographs of a tangential section of an adult photoreceptor cluster (**f**), and the same at higher magnification (**g**). **h**, Confocal section of a 55% p.d. eye stained for Dlt, Arm and phalloidin (Phl). **g**, **h**, Asterisk, photoreceptor; arrowhead, rhabdomere; arrow, rhabdomere stalk; double arrow, AJ. Scale bar, 1 μ m.

Fig. 2c), forming thick and highly irregular junctions. A top view of wild-type ommatidia shows that AJs form a tight circle (arrowhead, Fig. 2d; see also Fig. 1e) surrounding the apical domain of the photoreceptor cluster (asterisk, Fig. 2d). However, AJs in *crb*⁻ photoreceptors (arrow, Fig. 2d) stretch further basolaterally, leading to the appearance of disorganized patches of Arm staining throughout the lateral membrane. The AJ defects appear before the

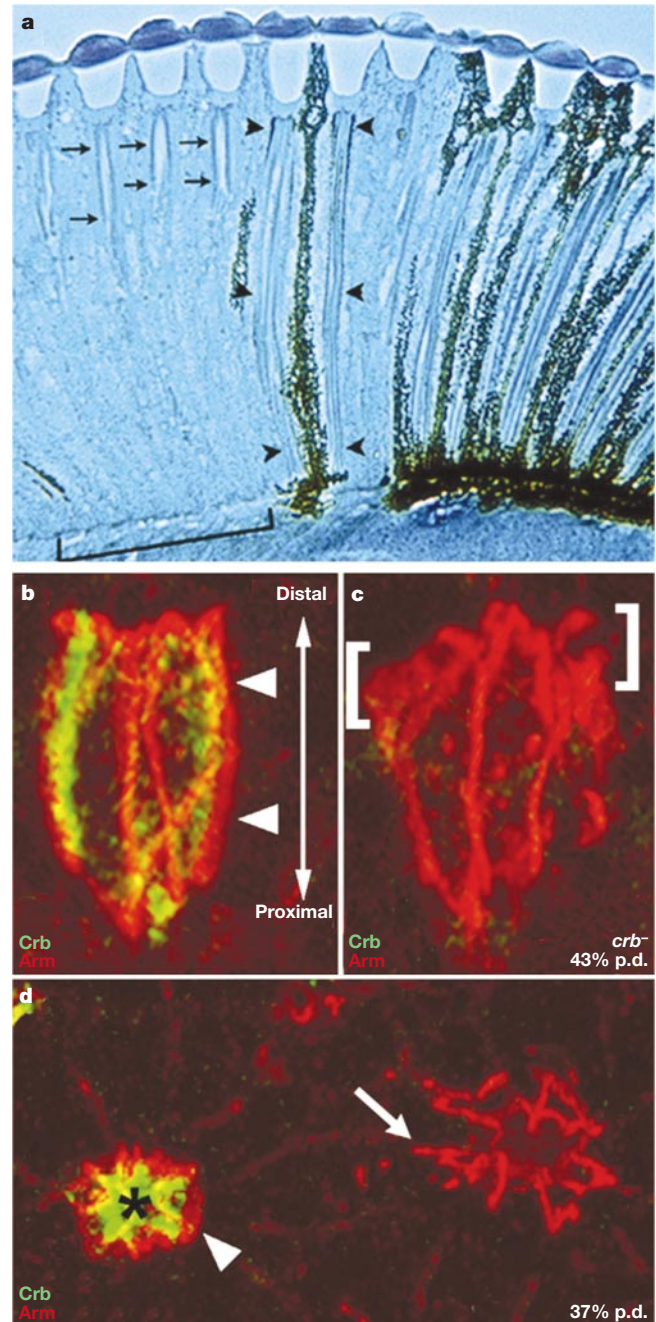


Figure 2 *crb*⁻ photoreceptors fail to extend rhabdomeres and accumulate Arm distally. **a**, Longitudinal section of a *crb*^{D88-3} adult eye clone containing wild-type (*w*⁺, arrowhead) and *crb*⁻ (*w*⁻, arrows) photoreceptors; bracket, relatively intact *w*⁺ retinal floor. Distal is to the top. **b–d**, 43% p.d. (**b**, **c**) and 37% p.d. (**d**) eyes containing *crb*^{11A22} clones marked by Crb and Arm. **b**, **c**, Longitudinal three-dimensional (3D) reconstruction of a wild-type (**b**) and a *crb*⁻ photoreceptor (**c**) in the same eye. Arrowheads, AJ; brackets, distal apical accumulation of Arm. **d**, Top view of a 3D reconstruction of a wild-type (left) and a *crb*⁻ (right) pupal photoreceptor. Black asterisk, apical Crb staining; arrowhead, AJ; arrow, Arm staining in *crb*⁻ cluster.

formation of the rhabdomere and localization of Crb and Dlt to the rhabdomere stalk.

To understand why *crb*⁻ rhabdomeres do not extend, we examined initial stages of rhabdomere formation and elongation. Cross-sections of the distal apical domain of *crb*⁻ photoreceptors depict rhabdomeres that are larger than the wild type (Fig. 3a, b; 95% of 80 photoreceptors scored). However, the rhabdomeres are very small at the proximal apical domain (Fig. 3a, d), even though cell membranes are present as marked by Dlt (Fig. 3a, c, e). Longitudinal

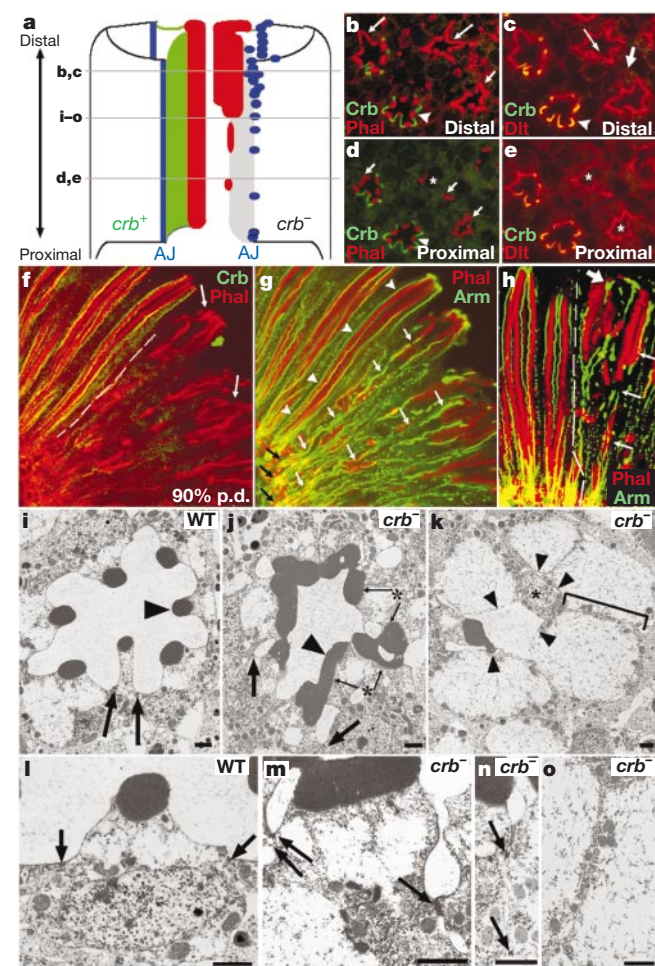


Figure 3 *crb*⁻ photoreceptor AJs and rhabdomeres are mispositioned without loss of polarity. **a**, Diagram of *crb*⁺ (left) and *crb*⁻ (right) photoreceptors depicting the position of tangential sections in this figure along the proximal–distal axis. Distal is to the top. *crb*⁻ photoreceptors accumulate rhabdomeres and AJs distally. **b–e**, Tangential section of *crb*^{11A22} photoreceptors at 65% p.d. stained with either rhodamine–phalloidin (**b** and **d**) or for Dlt (**c** and **e**) and Crb. **b** and **c** are the same section, showing expression of different markers, and are distal portion of the same clone in **d** and **e**. **b**, Distal portion of *crb*⁻ (arrows) and *crb*⁺ (arrowhead) photoreceptors. **c**, Thick arrow, peripheral Dlt; thin arrow, apical Dlt; arrowhead, wild-type Dlt localization. **d**, Small (arrows), missing (asterisk), and wild-type (arrowhead) rhabdomeres. **e**, Asterisks, Dlt in proximal membrane. **f–h**, *crb*^{D88-3} clone at 90% p.d. **f** and **g** are images from the same clone; **h** is a different clone. **f**, Crb marks wild-type ommatidia left of the white outline. Arrows, enlarged rhabdomeres. **g**, Arrows, discontinuous AJs; arrowheads, wild-type AJs; black arrows, retinal floor attachment. **h**, *crb*⁻ clone, right of the white outline (Crb staining not shown). Thick arrow, distal AJ; arrows, no proximal AJ. **i–o**, Transmission electron micrograph of wild-type (**i** and **l**) and *crb*⁻ photoreceptors (**j**, **k** and **m–o**). All panels are sections from the same eye at the mid-retinal level to show *crb*⁻ photoreceptors (see **a**). *crb*⁻ clones were located and oriented to section *crb*⁻ photoreceptors. **j–o**, Arrowheads, rhabdomere; arrows, AJ. **j**, Small arrows with asterisk mark photoreceptors with multiple rhabdomeres. **k**, Asterisk marks an unidentified cell. Scale bars, 1 μ m.

views of near-adult or adult stages show that individual *crb*⁻ photoreceptors extend down to different levels in the retina (Fig. 3f–h). AJs of *crb*⁻ photoreceptors are variable in size at the distal portion of the retina and discontinuous and thin in the proximal retina (arrows, Fig. 3g, h). The AJs attaching the photoreceptors to the floor of the retina remain relatively intact (Fig. 3g, h). Therefore, *crb*⁻ photoreceptor rhabdomeres and AJs are enlarged basolaterally in the distal domain of the photoreceptor and fail to extend proximally (Fig. 3a).

Transmission electron microscopy (TEM) was used to directly examine the structure of the *crb*⁻ photoreceptor rhabdomeres and AJs in adult eyes. Because *crb*⁻ photoreceptors contain rhabdomeres and AJs of variable length, tangential sections at the mid-retina level revealed both bulky rhabdomeres and small or no rhabdomeres (Fig. 3a, j and k, respectively). In addition, *crb*⁻ photoreceptors contain AJs of various sizes (Fig. 3a, m–o), ranging from large AJs (Fig. 3m) to no AJs (Fig. 3o). Rhabdomeres and AJs of wild-type clusters at the same level of the same mosaic eye are uniform in size and are properly positioned (Fig. 3a, i, l). *crb*⁻ photoreceptors retain apicobasal cell polarity because their rhabdomeres localize to the apex of the apical domain and their AJs are positioned laterally to the rhabdomeres. Enlarged rhabdomeres rarely overlap with AJs either by TEM (Fig. 3j–o) or by fluorescent imaging (Fig. 3g, h). Furthermore, although Dlt is not concentrated in the rhabdomere stalk, a significant amount remains apical, suggesting that the cell retains cell polarity. Therefore, Crb provides the positional cues that allow for extension of AJs and rhabdomeres along the growing proximal apical domain of the photoreceptor, independently of its role in determining apicobasal polarity.

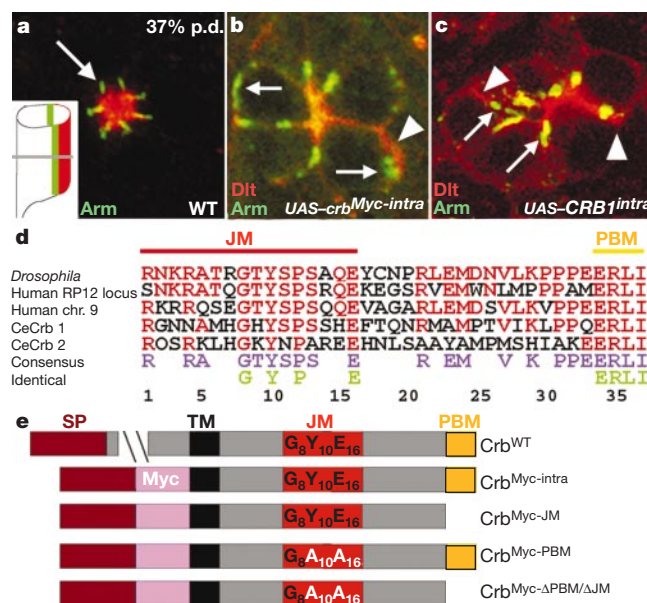


Figure 4 Overexpression of *Crb*^{Myc-intra} and human *CRB1*^{intra} mislocalizes AJs and Dlt, leading to loss of cell polarity. **a–c**, Tangential section of 37% p.d. wild-type (WT, **a**) and *pGMR-GAL4/UAS-crb*^{Myc-intra} (**b**) eyes grown at 18 °C, and *pGMR-GAL4/UAS-CRB1*^{intra} (**c**) grown at 25 °C. Arrows, AJs (ectopic in **b** and **c**); arrowheads, Dlt (mislocalized in **b** and **c**). Section location is shown in the schematic diagram. **d**, Alignment of conserved amino acids in the intracellular domain of Crb and homologues from humans and the nematode *Caenorhabditis elegans* (CeCrb 1 and 2). Identical amino acids between the *Drosophila* and other homologues are denoted in red. Purple, residues identical in three of five members; green, residues identical in all five members. JM, juxtamembrane (red) region; PBM, PDZ binding motif (yellow). The JM contains R1-G8-T9-Y10-E16, similar to the 12-amino-acid FERM binding site of glycophorin C (R1-G8-T9-T10-E13), β -neurexins (R1-G8-S9-T10-E13) and syndecans (K1-G8-S9-T10-E13)^{22,23}. **e**, Scheme of different Crb proteins used in this study. SP, signal peptide; TM, transmembrane domain; Myc, Myc-epitope tag. Mutant constructs are described in text.

Because loss of Crb caused the mislocalization of AJs, overexpression of Crb throughout the cells might lead to sequestration of AJs. Overexpression of full-length Crb (Crb^{WT})² causes the mislocalization of Arm, Dlt and Dlg, leading to loss of cell polarity in the third-instar eye disc, followed by pupal lethality, with escapers lacking eyes (data not shown). To find the region of Crb that might sequester AJs, the Crb intracellular domain ($\text{Crb}^{\text{Myc-intra}}$), which is identical to wild-type Crb in its ability to rescue crb^- embryos, was overexpressed in the eye². Overexpression of $\text{Crb}^{\text{Myc-intra}}$ in the eye caused a slightly milder phenotype than Crb^{WT} , as the adult eye is present but rough. Two phenotypes were observed in pupal eyes. First, 97% of the photoreceptors contained multiple ectopic AJ complexes of variable length ($n = 324$; Fig. 4b). Second, Dlt protein was mislocalized to other areas in the membrane (Fig. 4b), instead of its normal localization to the apical domain (Fig. 4a). Therefore, overexpression of $\text{Crb}^{\text{Myc-intra}}$ causes mislocalization of both photoreceptor AJs and Dlt throughout the cell membrane, resulting in loss of polarity.

Mutations in a human homologue of *crb*, *CRB1*, result in photoreceptor degeneration in RP12 and LCA⁶⁻⁸. We and others (ref. 7, and M. Pinto and U. Tepass, personal communication) have identified sequences in the *CRB1* locus homologous to the intracellular domain of *Drosophila* Crb (Fig. 4d) (these sequence data were produced by the Sanger Centre and can be obtained at <http://www.sanger.ac.uk> and from GenBank). To determine whether the properties of the Crb intracellular domain are conserved across species, we overexpressed the *CRB1* intracellular domain ($\text{CRB1}^{\text{intra}}$) in the eye. Overexpression of human $\text{CRB1}^{\text{intra}}$ resulted in a phenotype weaker but similar to $\text{Crb}^{\text{Myc-intra}}$, suggesting that

the function of the Crb intracellular domain is conserved across species (Fig. 4c). The intracellular domain of Crb contains two conserved regions: a juxtamembrane (JM) region similar to the band 4.1 binding site on glycophorin C, and a PDZ binding motif (PBM) at amino acids 34–37 at the carboxy terminus (Fig. 4d, e)^{4,15}. To define the portion of the Crb protein responsible for AJ positioning, we used versions of $\text{Crb}^{\text{Myc-intra}}$ containing the JM region but lacking the PDZ binding motif ($\text{UAS-cr}^{\text{Myc-JM}}$), or containing mutations replacing the conserved Y10 and E16 of the JM region with alanine but retaining a wild-type PDZ binding motif ($\text{UAS-cr}^{\text{Myc-PBM}}$) (Methods, Fig. 4d, e)⁴. As a control, we used $\text{UAS-cr}^{\text{Myc-}\Delta\text{PBM}/\Delta\text{JM}}$, in which all three mutations were present⁴. Flies overexpressing $\text{Crb}^{\text{Myc-}\Delta\text{PBM}/\Delta\text{JM}}$ in the eye had no significant phenotype (Fig. 5a, d, g), indicating that these two regions, JM and PBM, are responsible for the phenotypes associated with the overexpression of the Crb intracellular domain.

To determine which region of Crb was responsible for the ectopic AJ phenotype, we overexpressed $\text{Crb}^{\text{Myc-JM}}$. This causes mislocalization of AJs similar to that observed with $\text{Crb}^{\text{Myc-intra}}$ (compare arrows in Figs 5b and 4b). $\text{Crb}^{\text{Myc-JM}}$ pupal eyes exhibit lengthened or ectopic AJs in 70% of photoreceptors scored ($n = 280$; see Methods). However, Dlt protein localization was unaltered: Dlt localized properly to the centre of the cluster (Fig. 5b). Overexpression of $\text{Crb}^{\text{Myc-PBM}}$, which carries a wild-type PBM but mutations in the JM region, did not result in ectopic AJs (arrow, Fig. 5c), despite mislocalization of Dlt (arrowhead, Fig. 5c). Some 89% of photoreceptors contained punctate and single AJs ($n = 624$; arrow, Fig. 5c). A significant amount of Dlt remains at the apical pole of the cell, presumably bound to the endogenous Crb (Fig. 5c).

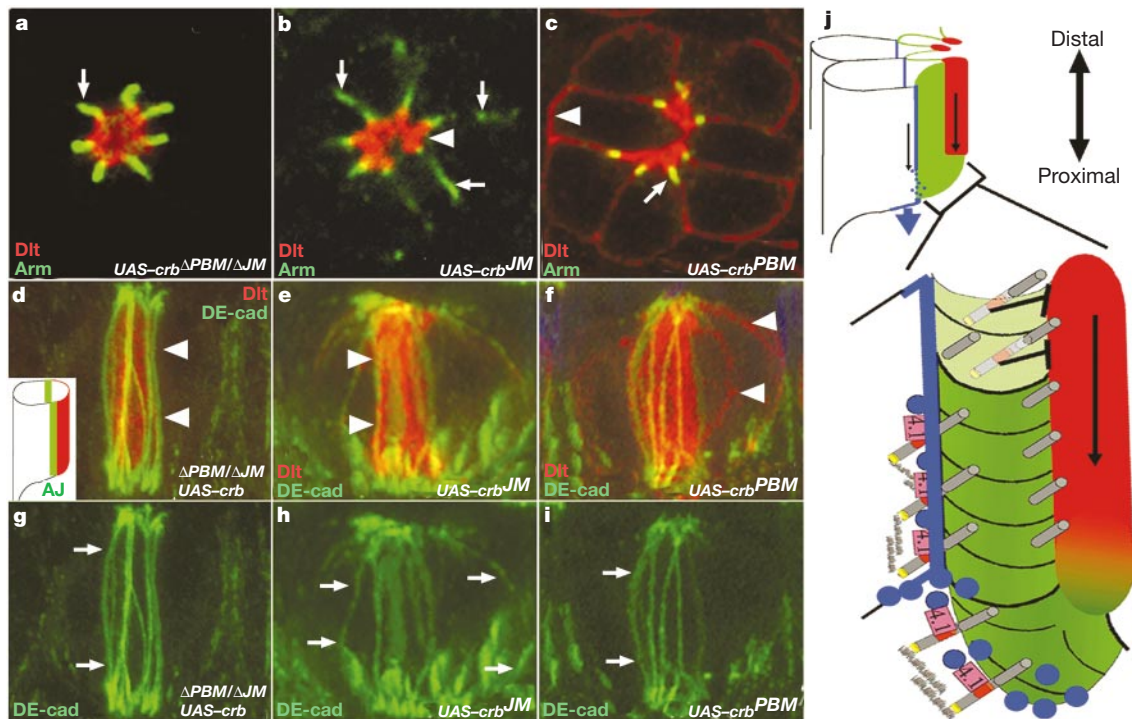


Figure 5 The juxtamembrane region of Crb is sufficient to recruit AJs independently of Dlt or polarity defects. In all panels, arrows indicate AJs and arrowheads mark Dlt localization. Panels **a–c** are tangential sections of an eye at 37% p.d. labelled for Dlt and Arm. **a**, $\text{pGMR-GAL4/UAS-cr}^{\text{Myc-}\Delta\text{PBM}/\Delta\text{JM}}$. **b**, $\text{pGMR-GAL4/UAS-cr}^{\text{Myc-JM}}$. **c**, $\text{pGMR-GAL4/UAS-cr}^{\text{Myc-PBM}}$. **d–i**, Three-dimensional reconstruction of longitudinal sections of photoreceptors (37% p.d.). **d**, $\text{pGMR-GAL4/UAS-cr}^{\text{Myc-}\Delta\text{PBM}/\Delta\text{JM}}$. **e**, $\text{pGMR-GAL4/UAS-cr}^{\text{Myc-JM}}$. **f**, $\text{pGMR-GAL4/UAS-cr}^{\text{Myc-PBM}}$. **g**, $\text{pGMR-GAL4/UAS-cr}^{\text{Myc-}\Delta\text{PBM}/\Delta\text{JM}}$. **h**, $\text{pGMR-GAL4/UAS-cr}^{\text{Myc-JM}}$. **i**, $\text{pGMR-GAL4/UAS-cr}^{\text{Myc-PBM}}$. **j**, Model of Crb action at the growing proximal apical domain of the developing photoreceptor. Rhabdomeres, red;

rhabdomere stalk, green; AJs, blue. As the retina deepens proximally (top), it pulls (thick blue arrow) the apical domain resulting in addition of membrane, AJ (blue) and rhabdomere (red). Crb proteins (bottom; grey cylinders) restrict the rhabdomere apically and force it to stretch proximally. Simultaneously, proximal Crb organizes AJ material (blue circles) into a continuous AJ through its FERM binding site (red box in Crb) that binds a putative FERM protein (pink boxes denoted by 4.1). The PBM of Crb (yellow box in Crb cylinder) binds Dlt (grey fuzzy molecule), which homomultimerizes¹⁵ to localize Crb to the rhabdomere stalk (green).

To determine whether $\text{Crb}^{\text{Myc-JM}}$ overexpression affects AJs, and not just Arm localization, we stained eyes overexpressing the three different $\text{Crb}^{\text{Myc-intra}}$ proteins for Dlt and *Drosophila* E-cadherin (DE-cad). Eyes overexpressing $\text{Crb}^{\text{Myc-}\Delta\text{PBM}/\Delta\text{JM}}$ depict wild-type Dlt, localized to the apical domain (Fig. 5d), and DE-cad staining, marking the AJs, which smoothly span the apical domain (Fig. 5g). Overexpression of $\text{Crb}^{\text{Myc-JM}}$ results in mislocalization of DE-cad throughout the cell membrane (arrows, Fig. 5h) while Dlt is retained at the most apical part of the photoreceptor cells (arrowhead, Fig. 5e). The retention of Dlt at the apical domain of the photoreceptors suggests that the cells retain apicobasal polarity. Conversely, overexpression of $\text{Crb}^{\text{Myc-PBM}}$ mislocalizes Dlt throughout the cell membrane (arrowheads in Fig. 5f) but DE-cad remains localized at the normal position of the AJ (Fig. 5i). Importantly, nearly all of the ommatidia in eyes overexpressing $\text{Crb}^{\text{Myc-PBM}}$ (97%, $n = 1,273$) and $\text{Crb}^{\text{Myc-JM}}$ (96%, $n = 830$) contain photoreceptors that retained cell polarity. Mislocalization of the Dlt throughout the membrane using the $\text{Crb}^{\text{Myc-PBM}}$ construct did not result in ectopic AJ formation, suggesting that the AJ defects seen in crb^- photoreceptors did not result from mislocalization of Dlt outside the rhabdomere stalk (Fig. 3). Thus, our data indicate that overexpression of the JM region of Crb recruits AJs to ectopic sites of the cell membrane without causing loss of apicobasal polarity.

Crb regulates cell polarity and AJ formation in the *Drosophila* embryo and follicular epithelium, but whether Crb controls AJs by regulating cell polarity or by an independent pathway is not known^{1,18,19}. Our work reveals a role for Crb in localizing photoreceptor AJs and rhabdomeres, independent of apicobasal polarity. Furthermore, we show that the putative FERM binding site of Crb controls AJ localization, which may act either through a FERM protein or spectrins (modelled in Fig. 5j).

Various mutations in the human homologue of *crb*, *CRB1*, cause two severe, early-onset photoreceptor-degeneration diseases, RP12 and LCA⁶⁻⁸. LCA patients seem to carry more severe mutations at the *CRB1* locus and are blind from early infancy^{9,10}. The photoreceptors in these patients, although present, have short or absent inner and outer segments^{9,10}, analogous to the short crb^- photoreceptors in *Drosophila*. Furthermore, the *CRB1* intracellular domain seems functionally conserved in fly eyes (Fig. 4c). Thus, studies of Crb in *Drosophila* eyes may help to unravel the pathogenesis of related human diseases. □

Methods

Fly strains

The isolation of $\text{crb}^{\text{D88-3}}$ and crb^{1A22} has been previously described¹ as have other mutant strains²⁰. *ey-FLP* (provided by B. Dickson) and *FRT82B* were used for FRT/FLP mosaic analysis^{16,17}. The *UAS/GAL4* system was used for overexpression experiments²¹. Crb transgenic lines including *UAS-crb^{WT}*, *UAS-crb^{Myc-intra}* (ref. 2), *UAS-crb^{Myc-intra} $\Delta\text{PBM}/\Delta\text{JM}$* , *UAS-crb^{Myc-JM}* and *UAS-crb^{Myc-PBM}* (also known as *UAS-crb^{Myc-intra}Y10A/E16A/ ΔERLI* , *UAS-crb^{Myc-intra}Y10A/E16A* and *UAS-crb^{Myc-intra} ΔERLI* in ref. 4), and were donated by E. Knust. *pGMR-GAL4* was supplied by M. Freeman.

Induction of mosaic clones and ectopic Crb

Clones of $\text{crb}^{\text{D88-3}}$ and crb^{1A22} were generated using *ey-FLP;FRT82B crb^{D88-3}/FRT82B* and *ey-FLP;FRT82B crb^{1A22}/FRT82B*, respectively. For adult dissections, clones were induced in a *w⁻* background. The C-terminal part, including the intracellular and transmembrane domain, of human *CRB1* was amplified by polymerase chain reaction (PCR) from a human brain complementary DNA library (Clontech) using the primers 5'-TTCGAGATCTTTCACCACTATTGGCTCAGTGAC-3' and 5'-TGTGCTCGAGCATCTCGAAGGGACACATGCTC-3'. This fragment was fused with the signal peptide sequence of fly *crb*, and inserted into *pUAST*, using a previously described method². Different mutant forms of Crb protein were overexpressed by using *pGMR-GAL4*, which is expressed in differentiating retinal cells. $\text{Crb}^{\text{Myc-intra}}$ was overexpressed in flies kept at 18 °C and $\text{CRB1}^{\text{intra}}$ in flies kept at 25 °C, whereas all other constructs were overexpressed in flies kept at room temperature. At least five independent eyes were used to score for defective photoreceptor AJ in both *pGMR-GAL4/UAS-crb^{Myc-JM}* and *pGMR-GAL4/UAS-crb^{Myc-PBM}*, whereas three independent eyes were used to score abnormal contacts between photoreceptors.

Histology, immunohistochemistry and confocal microscopy

TEM was done as previously described¹³, except 1.25% glutaraldehyde and 1% paraformaldehyde in 0.1 M sodium phosphate buffer was used as fixative. Eyes from pupae aged at room temperature were prepared as previously described¹³ and fixed in 2% paraformaldehyde-lysine-periodate (PLP). DE-cad staining was carried out in solutions containing 1 mM CaCl_2 . Late-pupal-stage eyes stained for Crb and phalloidin were fixed as above and then placed in ice-cold acetone for 10 min. All primary antibodies used were rabbit anti-Dlt (1:500), mouse anti-Dlt (1:500), rat anti-Crb (1:400), mouse anti-Arm (1:200; from M. Peifer), guinea pig anti-Dlg (1:1000; from P. Bryant), rabbit anti-Dlg (1:500; from K.-O. Cho), and rat anti-DE-cad (1:50; from M. Takeichi). Fluorescein isothiocyanate (FITC)-conjugated phalloidin was from Sigma, whereas fluorescent secondary antibodies were from Jackson Immunochemicals. Images were scanned using a Zeiss LSM laser-scanning confocal microscope.

Received 30 November 2001; accepted 22 January 2002.

Published online 17 February 2002, DOI 10.1038/nature720.

1. Tepass, U., Theres, C. & Knust, E. *crumbs* encodes an EGF-like protein expressed on apical membranes of *Drosophila* epithelial cells and required for organization of epithelia. *Cell* **61**, 787–799 (1990).
2. Wodarz, A., Hinz, U., Engelbert, M. & Knust, E. Expression of crumbs confers apical character on plasma membrane domains of ectodermal epithelia of *Drosophila*. *Cell* **82**, 67–76 (1995).
3. Wolff, T. & Ready, D. F. in *The Development of Drosophila melanogaster* (ed. Bate, M. Martinez-Arias, A.) Cold Spring Harbor Press, Cold Spring Harbor, New York, (1993), 1277–1325.
4. Klebes, A. & Knust, E. A conserved motif in the cytoplasmic domain of *Drosophila* Crumbs is required for DE-cadherin localisation and zonula adherens formation. *Curr. Biol.* **10**, 76–85 (2000).
5. Chishti, A. H. *et al.* The FERM domain: a unique module involved in the linkage of cytoplasmic proteins to the membrane. *Trends Biochem. Sci.* **23**, 281–282 (1998).
6. den Hollander, A. I. *et al.* Mutations in a human homologue of *Drosophila* crumbs cause retinitis pigmentosa (RP12). *Nature Genet.* **23**, 217–222 (1999).
7. den Hollander, A. I. *et al.* Leber congenital amaurosis and retinitis pigmentosa with Coats-like exudative vasculopathy are associated with mutations in the *crumbs* homologue 1 (*CRB1*) gene. *Am. J. Hum. Genet.* **69**, 198–203 (2001).
8. Lotery, A. J. *et al.* Mutations in the *CRB1* gene cause Leber congenital amaurosis. *Arch. Ophthalmol.* **119**, 426–427 (2001).
9. Mizuno, K. *et al.* Leber's congenital amaurosis. *Am. J. Ophthalmol.* **83**, 32–42 (1977).
10. Noble, K. G. & Carr, R. E. Leber's congenital amaurosis. A retrospective study of 33 cases and a histopathological study of one case. *Arch. Ophthalmol.* **96**, 818–821 (1978).
11. Cagan, R. L. & Ready, D. F. The emergence of order in the *Drosophila* pupal retina. *Dev. Biol.* **136**, 346–362 (1989).
12. Chang, H. Y. & Ready, D. F. Rescue of photoreceptor degeneration in rhodopsin-null *Drosophila* mutants by activated Rac. *Science* **290**, 1978–1980 (2000).
13. Kumar, J. P. & Ready, D. F. Rhodopsin plays an essential role in *Drosophila* photoreceptor development. *Development* **121**, 4359–4370 (1995).
14. Longley, R. L. & Ready, D. F. Integrins and the development of three-dimensional structure in the *Drosophila* compound eye. *Dev. Biol.* **171**, 415–433 (1995).
15. Bhat, M. *et al.* Discs Lost, a novel multi-PDZ domain protein, establishes and maintains epithelial polarity. *Cell* **96**, 833–845 (1999).
16. Xu, T. & Rubin, G. M. Analysis of genetic mosaics in developing and adult *Drosophila* tissues. *Development* **117**, 1223–1237 (1993).
17. Golik, K. G. & Linquist, S. The FLP recombinase of yeast catalyzes site-specific recombination in the *Drosophila* genome. *Cell* **59**, 499–509 (1989).
18. Tepass, U. Crumbs, a component of the apical membrane, is required for zonula adherens formation in primary epithelia of *Drosophila*. *Dev. Biol.* **177**, 217–225 (1996).
19. Tanentzapf, G., Smith, C., McGlade, J. & Tepass, U. Apical, lateral and basal polarization cues contribute to the development of the follicular epithelium during *Drosophila* oogenesis. *J. Cell Biol.* **151**, 891–904 (2000).
20. Lindsley, D. L. & Zimm, G. *The Genome of Drosophila melanogaster* (Academic, San Diego, 1992).
21. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
22. Littleton, J. T., Bhat, M. A. & Bellen, H. J. Deciphering the function of Neurexins at cellular junctions. *J. Cell Biol.* **137**, 793–796 (1997).
23. Rapraeger, A. C. & Ott, V. L. Molecular interactions of syndecan core proteins. *Curr. Opin. Cell Biol.* **10**, 620–628 (1998).

Acknowledgements

We especially thank K.-O. Cho for her contribution to experimental design and for her support. We thank E. Knust for reagents and U. Tepass for discussion of unpublished results. We thank J. Chern and J. Lim for comments and discussion, B. Kehl and Z. H. Chen for technical assistance. We thank the Indiana Stock Center for providing flies. We thank H. Adams for technical assistance with TEM. Confocal microscopy was supported by a grant from the National Institutes of Health to D. B. Jones. S.I. and S.-C.N. are postdoctoral fellows. H.J.B. is a Howard Hughes Medical Institute Investigator. M.A.B. is supported by a Howard Temin Career Development Award from the National Cancer Institute. M.A.B., H.J.B., and K.-W.C. are supported by the NIH and K.-W.C. by the Retina Research Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.-W.C. (e-mail: kchoi@bcm.tmc.edu).

Cbl–CIN85–endophilin complex mediates ligand-induced downregulation of EGF receptors

Philippe Soubeyran^{*†}, Katarzyna Kowanetz^{*†}, Iwona Szymkiewicz^{*†}, Wallace Y. Langdon[‡] & Ivan Dikic^{*}

^{*} Ludwig Institute for Cancer Research, Box 595, Husargatan 3, Uppsala, S-75124, Sweden

[‡] Department of Pathology, University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia

[†] These authors contributed equally to this work

Cbl is a multi-adaptor protein involved in ligand-induced downregulation of receptor tyrosine kinases. It is thought that Cbl-mediated ubiquitination of active receptors is essential for receptor degradation and cessation of receptor-induced signal transduction^{1–5}. Here we demonstrate that Cbl additionally regulates epidermal growth factor (EGF) receptor endocytosis. Cbl rapidly recruits CIN85 (Cbl-interacting protein of 85K; ref. 6) and endophilins (regulatory components of clathrin-coated vesicles^{7–10}) to form a complex with activated EGF receptors, thus controlling receptor internalization. CIN85 was constitutively associated with endophilins, whereas CIN85 binding to the distal carboxy terminus of Cbl was increased on EGF stimulation. Inhibition of these interactions was sufficient to block EGF receptor internalization, delay receptor degradation and enhance EGF-induced gene transcription, without perturbing Cbl-directed receptor ubiquitination. Thus, the evolutionary divergent C terminus of Cbl uses a mechanism that is functionally separable from the ubiquitin ligase activity of Cbl to mediate ligand-dependent downregulation of receptor tyrosine kinases.

Recent genetic evidence has indicated that the C terminus of SLI-1, the Cbl orthologue in *Caenorhabditis elegans*, is necessary for negative regulation of LET-23-induced vulval development in nematode worms¹¹. The C terminus of Cbl is divergent from SLI-1, and contains a larger proline-rich region and a distal C-terminal tail composed of an acidic box domain and a ubiquitin-associated (UBA) domain¹. To identify proteins that bind specifically to the C terminus of Cbl, we screened for clones interacting with either the proline-rich sequences (residues 480–680) or the distal part (residues 680–906) of Cbl in the yeast two-hybrid system. Among six double positive clones, one clone coding for CIN85 bound to the distal tail of Cbl (Supplementary Information Fig. 1a). CIN85, also known as Ruk¹² and SETA¹³, is an adapter protein containing three Src-homology 3 (SH3) domains, a proline-rich region and a coiled-coil domain⁶. The SH3 domains of CIN85 bound to the distal C terminus of Cbl but not to the proline-rich region of Cbl in pull-down assays (Supplementary Information Fig. 1b, c), and in co-precipitation studies in mammalian cells (Fig. 1a). We also observed increased complex formation between CIN85 and Cbl in EGF-stimulated HeLa and Chinese hamster ovary (CHO) cells (Fig. 1b), as well as in human embryonic kidney (HEK) 293T cells transfected with Cbl and CIN85 constructs (Supplementary Information Fig. 1d). CIN85 binding to Cbl was largely attenuated in a phosphorylation-defective Cbl mutant (Y700/731/774F) (data not shown and ref. 6). Similarly, Cbl association with CMS, a homologue of CIN85, is regulated by tyrosine phosphorylation of Cbl¹⁴. Therefore, EGF-induced phosphorylation of Cbl may cause a conformational change in its distal C terminus, thus exposing polyproline motifs that serve as binding sites for the SH3 domains of CIN85 or CMS. Notably, a ligand-inducible complex between EGF receptor and CIN85 was detected in cells expressing Cbl or oncogenic Cbl with impaired RING-finger domain (Cbl-70Z), but

not Cbl deleted for CIN85 binding sites (Cbl-655) (Fig. 1c). These data indicate that increased Cbl phosphorylation and binding to EGF receptors, but not its ubiquitin ligase activity, contribute to the recruitment of CIN85 in EGF receptor complexes.

CIN85 was also shown to associate with several adaptor proteins^{12,15,16}, and could thereby facilitate the assembly of a larger Cbl-associated network on growth-factor stimulation. To identify additional CIN85-associated proteins, we screened for clones able to interact with the C terminus of CIN85 in the yeast two-hybrid system. Seven double positive clones were identified as endophilin A1, A2 or A3, each containing the intact SH3 domain (not shown), suggesting that the SH3 domain of endophilins binds to CIN85. These results were confirmed *in vitro* by the ability of the SH3 domain of endophilin A1 to bind to the isolated proline-rich

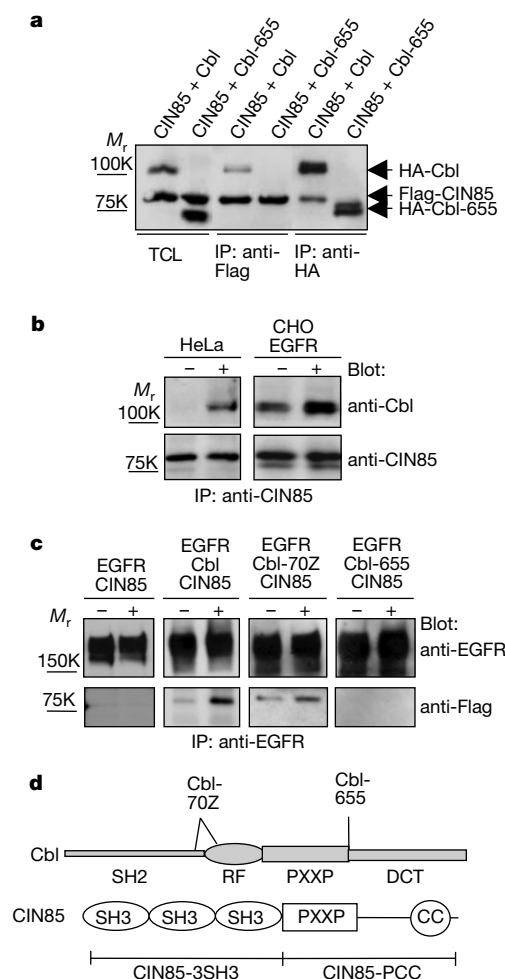


Figure 1 CIN85 is recruited to Cbl–EGF receptor complexes on EGF stimulation.

a, Lysates (TCL) of HEK 293T cells expressing indicated proteins were immunoprecipitated (IP) with anti-HA or anti-Flag antibodies, and immunoblotted with a mixture of anti-HA/Flag antibodies. *M_r*, relative molecular mass. **b**, Lysates of unstimulated (–) or stimulated (EGF 50 ng ml^{–1} for 5 min at 37 °C) cells were subjected to immunoprecipitation with an antibody against CIN85 (CT), and blotted with anti-Cbl or anti-CIN85 (CT) antibodies. **c**, Lysates of HEK 293T cells expressing designated proteins were subjected to immunoprecipitation with anti-EGFR and immunoblotting with anti-EGFR or anti-Flag (M5) antibodies. Anti-Flag antibodies recognized Flag-tagged CIN85. **d**, Structure of Cbl and CIN85 and schematic representation of constructs used in these studies. SH2, Src homology 2 domain; RF, RING-finger ubiquitin ligase domain; PXXP, polyproline region; DCT, distal C terminus of Cbl; 3SH3, three SH3 domains; CC, coiled-coil. Cbl-655 is a deletion of the DCT of Cbl; Cbl-70Z is an internal deletion of 17 amino-acid residues in the beginning of the RING finger.

sequence of CIN85, and vice versa (Supplementary Information Fig. 2a, b). In addition, endogenous or overexpressed Cbl co-precipitated with endophilin A1 only in the presence of CIN85, indicating that CIN85 acts as a linker between Cbl and endophilins (Fig. 2a). CIN85 was constitutively bound to endophilin A1 (Fig. 2b), whereas its binding to Cbl was increased in EGF-stimulated cells (Figs 1b and 2b), thereby leading to EGF-induced co-precipitation between Cbl and endophilin A1 (Fig. 2b) and EGF receptors and endophilin A1 (Fig. 2c). Importantly, we were able to demonstrate a ligand-inducible association of endogenous CIN85 and endophilins with the Cbl-EGFR complex in HeLa and MDA cells (Fig. 2d). CIN85 and endophilins co-precipitated with—in addition to phosphorylated EGF receptors and Cbl—several unidentified tyrosine phosphorylated proteins from EGF-treated cells (Fig. 2d). Taken together our data indicate that CIN85 constitutively associates with endophilins, whereas its potency to bind to Cbl is increased on EGF stimulation. In this way, CIN85 recruits endophilins in the complex with activated EGF receptors.

Endophilins were previously implicated in the control of clathrin-mediated endocytosis^{7–10}. They bind to the regulatory components of endocytic vesicles^{17,18} and modulate plasma membrane invagination by intrinsic acyl-transferase activity⁷ and by direct binding to lipid bilayers¹⁹. Both CIN85 and endophilins were found diffusely distributed in the cytoplasm of resting CHO cells, whereas

they co-localized with EGF receptors in endocytic vesicles after EGF stimulation (Supplementary Information Fig. 3a, b). We next tested the importance of a Cbl–CIN85–endophilin complex for EGF receptor ubiquitination and endocytosis. Dominant interfering forms of CIN85 containing either isolated SH3 domains (CIN85-3SH3) that associate with Cbl (Supplementary Information Fig. 1c) or a proline-rich region with a coiled-coil domain (CIN85-PCC) that binds to endophilin A1 (Supplementary Information Fig. 2a, b), were analysed for their ability to modulate EGF receptor polyubiquitination and subsequent receptor degradation. Cbl-mediated polyubiquitination of the EGF receptor was intact in the presence of dominant interfering forms of CIN85 (Fig. 3a). In contrast, the expression of CIN85-3SH3 or CIN85-PCC, but not CIN85, delayed Cbl-mediated EGF receptor degradation (Fig. 3b). These data suggest that CIN85 is dispensable for polyubiquitination of EGF receptors, whereas it may be critical for the regulation of receptor endocytosis and lysosomal degradation. Furthermore, these results indicate that polyubiquitinated EGF receptors can be uncoupled from a degradative fate if parallel Cbl/CIN85-mediated events are perturbed.

We investigated further the impact of expression of dominant interfering mutants of CIN85 on EGF receptor endocytosis and downregulation. Expression of Cbl together with EGF receptors led to a significant increase in the rate of ligand internalization and the

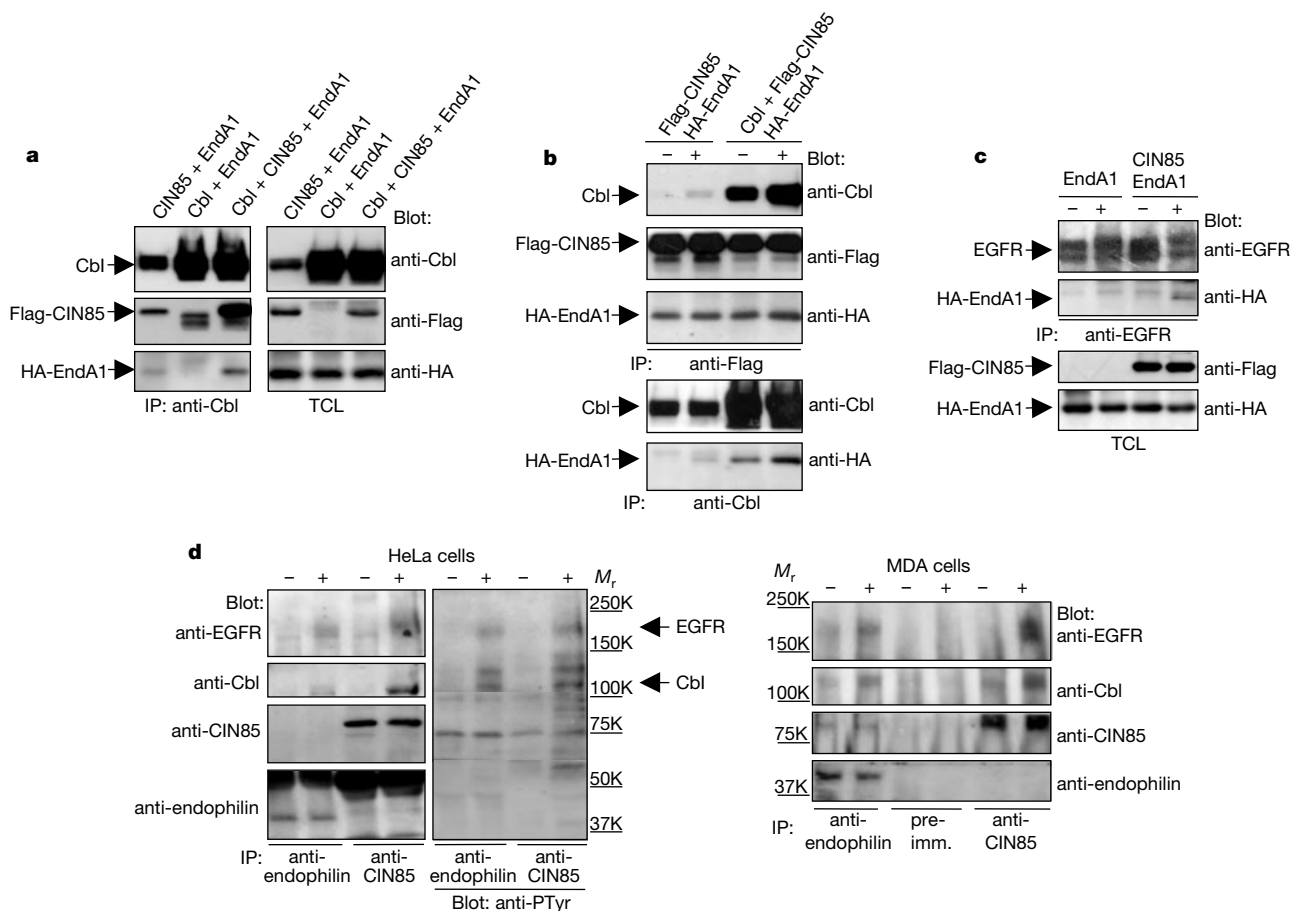


Figure 2 CIN85 links endophilins with the Cbl–EGFR complex. **a**, Lysates (TCL) of HEK 293T cells expressing designated proteins were subjected to immunoprecipitation (IP) and immunoblotting with indicated antibodies. EndA1, endophilin A1. **b**, **c**, HEK 293T cells expressing indicated proteins were left unstimulated (–) or stimulated with EGF (50 ng ml^{–1}) for 5 min at 37 °C (+). Lysates were immunoprecipitated with anti-Flag (M2), anti-Cbl (RF) and anti-EGFR antibodies, and blotted with corresponding antibodies. **d**, HeLa and MDA cells were starved (–) or stimulated with EGF (50 ng ml^{–1}) for 5 min at

37 °C (+), and lysates were subjected to immunoprecipitation with antibodies against endophilins (mixed S-20 and P26), an anti-CIN85 (CT) serum or preimmune serum, followed by immunoblotting with the indicated antibodies. Co-precipitation between CIN85 and endophilins was observed only after long exposures owing to low immunoblotting sensitivity of anti-endophilin and anti-CIN85 antibodies. Membranes were stripped and reblotted with anti-phosphotyrosine (PY99) antibodies.

clearance of EGF receptors from the cell surface when compared with cells transfected with EGF receptors alone (Fig. 3c, d). However, Cbl-accelerated receptor internalization and downregulation was efficiently blocked by expression of CIN85-3SH3 and CIN85-PCC (Fig. 3c, d), but not by expression of CIN85 (Fig. 3c), single SH3 domains of CIN85 or polyproline motifs that do not bind to endophilins (Supplementary Information Fig. 4). The principal function of CIN85 was further demonstrated by the inhibitory effect of CIN85-3SH3 or CIN85-PCC on the endogenous machinery involved in EGF receptor internalization and downregulation in CHO cells (Fig. 3c, d). These inhibitory effects of CIN85-3SH3 and CIN85-PCC were also visualized by immunofluorescence (Fig. 3e), quantified and expressed as a percentage of CHO cells internalizing EGF receptors (Supplementary Information Fig. 3c). It was recently demonstrated that Cbl binds and ubiquitinates EGF receptors at the plasma membrane, and remains associated with EGF receptors along the endocytic pathway where it may participate in sorting internalized receptors for lysosomal degradation^{20,21}. Our results indicate that Cbl, by recruiting

CIN85 and endophilins in complex with EGF receptors, also has an important function in the regulation of receptor internalization.

Finally, we tested whether dominant interfering forms of CIN85 can attenuate the inhibitory effect of Cbl on EGF receptor signalling pathways, by measuring EGF-induced transcription of the serum response element (SRE). Expression of Cbl decreased—in a dose-dependent manner—transcription of SRE triggered by EGF (Fig. 4a), whereas co-expression of CIN85-3SH3 or CIN85-PCC, but not CIN85, led to a reduction in the ability of Cbl to block EGF-induced SRE activity (Fig. 4a). Therefore, the full inhibitory effect of Cbl on EGF receptor signalling depends on its interactions with a functional CIN85 protein. The fact that CIN85-PCC, which does not bind to Cbl, was also sufficient to reduce the inhibitory effects of Cbl supports the idea that CIN85-PCC interacting proteins mediate, in part, the inhibitory effects of Cbl on EGF receptor signalling.

Identification of ligand-inducible interactions between the EGF receptors and a CIN85–endophilin complex has shed new light on the mechanisms involved in the regulation of receptor endocytosis, as well as receptor-mediated signalling^{4,5}. Here we have defined a

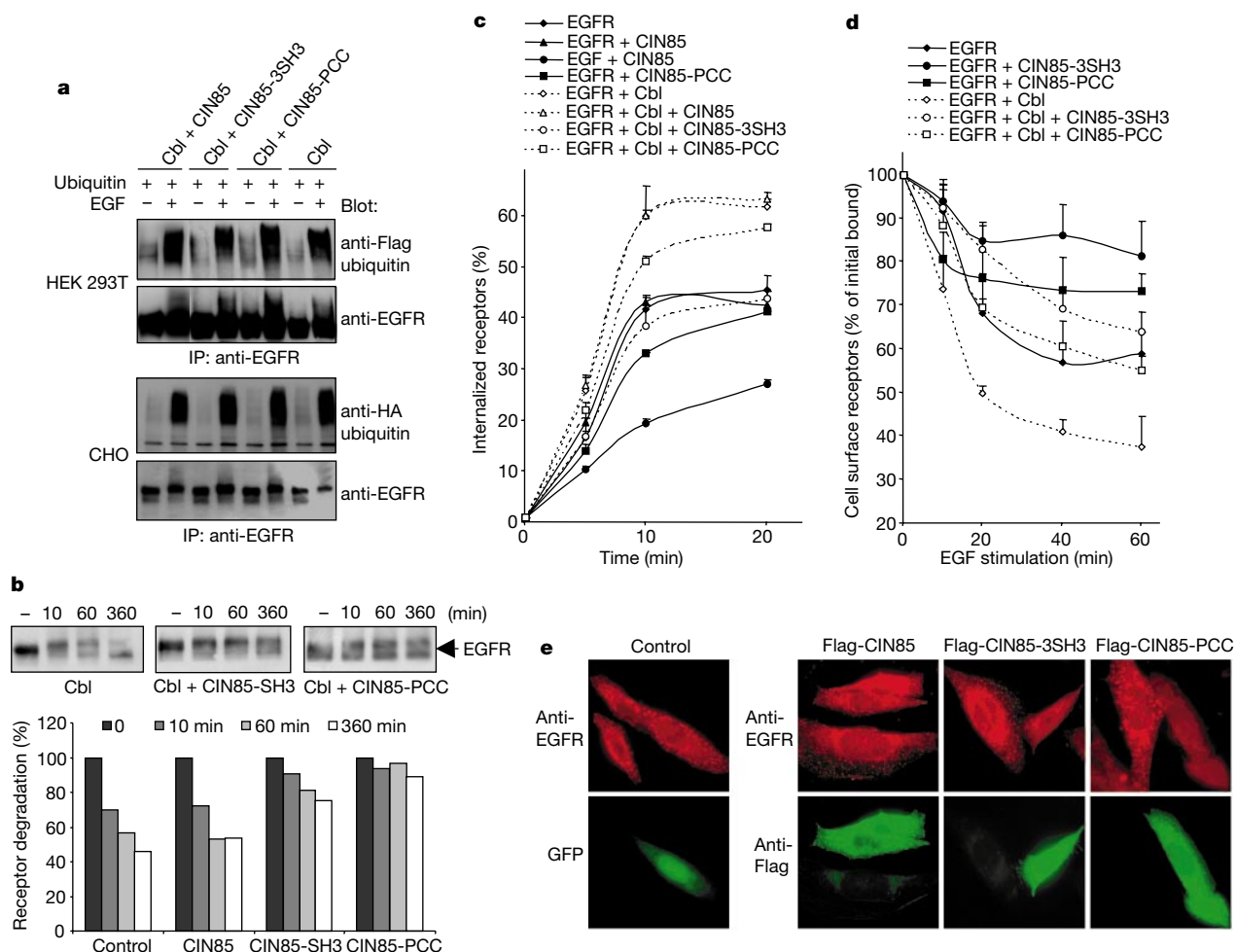


Figure 3 Effects of dominant interfering forms of CIN85 on Cbl-mediated ubiquitination, endocytosis and degradation of EGF receptors. **a**, HEK 293T or CHO cells were transfected with plasmids encoding EGFR and Flag- or HA-tagged ubiquitin in addition to Cbl alone or Cbl with CIN85, CIN85-3SH3 or CIN85-PCC constructs. Cells were treated with EGF (50 ng ml⁻¹) for 10 min at 37 °C (+), and immunoprecipitates of EGFR were blotted with anti-Flag, anti-HA or anti-EGFR antibodies. **b**, HEK 293T cells expressing EGFR and designated proteins were stimulated for the indicated time periods, and lysates were subjected to immunoblotting with anti-EGFR antibodies (upper panel). The levels of EGFR in cells were quantified and expressed as a percentage of remaining receptors for each time point. **c**, Ligand internalization in CHO cells transfected with EGFR alone or together

with the indicated proteins was performed as described in Methods. The results are expressed as a percentage between internalized and total cell-associated radioactive EGF. **d**, The level of surface receptors following time-lapse EGF stimulation was measured by a downregulation assay. The results are expressed as a percentage of the ¹²⁵I-labelled EGF bound to cell surface receptor after stimulation by non-labelled EGF for the indicated times. **e**, CHO EGFR cells expressing the indicated proteins were subjected to immunofluorescence analysis using anti-EGFR and anti-Flag (M2), followed by anti-rabbit TRITC (tetramethylrhodamine isomer R) and anti-mouse fluorescein isothiocyanate antibodies. A representative picture of the inhibition of receptor internalization and vesicle formation in cells expressing CIN85-3SH3 and CIN85-PCC is shown. GFP, green fluorescent protein.

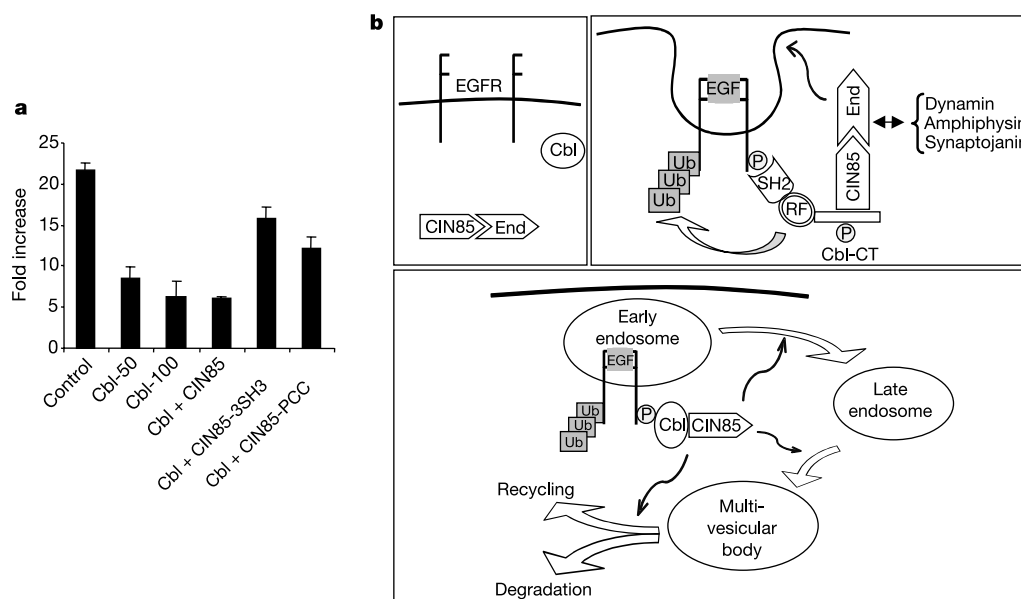


Figure 4 Role of CIN85 in the regulation of EGF-induced gene transcription. **a**, HEK 293T cells were transfected with reporter genes *SRE-Luc* (10 ng) and *lacZ* (20 ng), together with pRK5-EGFR (10 ng) alone (control) or with increasing amounts of Cbl (50 ng or 100 ng). In addition, vectors coding for CIN85 (100 ng), CIN85-3SH3 (100 ng) or CIN85-PCC (100 ng) were co-transfected with Cbl (100 ng). EGF-induced increase in luciferase activity was quantified and expressed as average $\pm$ s.d. **b**, Schematic model of EGF receptor downregulation. In resting cells (top left), CIN85 and endophilins (End) are constitutively associated and are not bound to Cbl or EGF receptors. Upon ligand-induced activation of EGF receptors (top right), Cbl binds to phosphorylated receptors and promotes receptor

ubiquitination (Ub). Cbl is also tyrosine phosphorylated in this complex, leading to translocation of CIN85/endophilin in the vicinity of active EGF receptors, whereby endophilins together with dynamin, amphiphysin and synaptojanin regulate clathrin-mediated endocytosis. Cbl is associated with EGF receptors along the endocytic pathway, and is implicated in sorting internalized receptors for lysosomal degradation (bottom). CIN85 may control EGF receptor trafficking by its ability to bind to multiple adaptor proteins, and thus promote the assembly of Cbl-linked signalling networks involved in cross-talk between endosome and the actin cytoskeleton.

regulatory role for Cbl, which is functionally separable from its ubiquitin ligase activity, and involves the recruitment of CIN85–endophilin complexes to activated EGF receptors (Fig. 4b). Translocation of endophilins in the vicinity of active EGF receptors promotes membrane invagination and thus receptor internalization, by acting together with other regulatory components of endocytic vesicle formation, including the GTPase dynamin^{7,9,17}. Accordingly, it has been demonstrated that overexpression of a dominant interfering form of endophilin prevents c-Met receptor endocytosis and enhances receptor-mediated biological responses²². The present study has also indicated a role for CIN85 in the control of post-membrane events such as targeting receptors for degradation and regulation of gene transcription (Figs 3b and 4a). CIN85 may control these processes by binding to multiple adaptor proteins (for example, Crk¹⁵, p130Cas¹⁵ and the p85 subunit of phosphatidylinositol-3 kinase¹²) involved in cross-talk between endosome and the actin cytoskeleton. The ability of CIN85 to assemble larger Cbl-associated signalling networks places it in a position to coordinate multiple steps in the endocytosis of tyrosine kinase receptors. Defining how this process is regulated in living cells will be the challenge for future investigations. □

Methods

Materials and plasmids

EGF was purchased from Intergen and ¹²⁵I-labelled EGF was from Amersham Biosciences. pSR-HA-Cbl, pSR-HA-Cbl-70Z, pSR-HA-Cbl-655 and pSR-HA-Cbl-480 or pRK5-Cbl were described previously^{23,24}. Haemagglutinin (HA)-tagged endophilin A1 in pcDNA3, Flag-tagged CIN85 in pcDNA3, Flag-tagged ubiquitin in pcDNA3.1, and pYTH9 vector was provided by P. De Camilli, S. Kajigaya, K. Tanaka and P. Aspenström, respectively. Description of CIN85 constructs is provided in Supplementary Information. CHO, MDA and HeLa cells were purchased from American Type Culture Collection; CHO cells stably expressing EGFR were provided by J. Borst and A. de Melker.

Yeast two-hybrid system screening

Yeast screening was performed using the Gal4-based Matchmaker two-hybrid system

with the human fetal brain, thymus and T-cell libraries (Clontech). The yeast retransformation and the filter lift assay were done as described in the Matchmaker system manual.

GST binding assays and immunochemical analyses

Glutathione S-transferase (GST) binding assays, immunoprecipitation and immunoblotting were performed as described^{25,26}. The following antibodies were used: anti-HA (12CA5; Roche), anti-Flag (M2 and M5; Sigma), mouse monoclonal anti-Cbl (C40320; Transduction Laboratories), rabbit polyclonal (RK2) anti-EGFR antibodies (a gift of J. Schlessinger), anti-endophilin A1 (S-20; Santa Cruz), pan-endophilin antibodies (Nuts) (a gift of P. De Camilli), anti-endophilin (P26) antibodies (a gift of N. Migone), rabbit polyclonal anti-Cbl antibodies raised against the GST fusion protein containing the RING finger of Cbl (RF), and rabbit polyclonal anti-CIN85 (CT) antibodies raised against the last 21 amino acids of human CIN85. Immunofluorescence studies were carried out as described²⁶, with specific details provided in Supplementary Information.

Ligand internalization and receptor downregulation assays

Ligand internalization and downregulation assays were performed as reported^{21,27}. Briefly, cell monolayers of transfected CHO cells were incubated for 1 h at 4 °C with ¹²⁵I-labelled EGF, washed twice with binding buffers, and then incubated at 37 °C for the indicated time intervals. Cells were transferred on ice and washed with either cold binding buffer or mild acidic buffer to remove surface-bound radiolabelled EGF. We quantified the remaining radioactivity in cells after cell lysis. Each point was measured in quadruplicate and expressed as a percentage (average $\pm$ s.d.) of internalized versus total cell-associated radioactive EGF. To measure EGF receptor downregulation, CHO cells were transfected in triplicates in 24-well plates, and 48 h later cells were either left unstimulated or incubated with EGF (50 ng ml⁻¹) at 37 °C for the indicated time intervals, subjected to acid wash to remove bound EGF, and subsequently incubated with ¹²⁵I-labelled EGF for 1.5 h at 4 °C. The amount of surface-bound radioactive EGF was determined by a γ -counter and expressed as a percentage of surface-labelled receptors compared with unstimulated cells. Details of receptor ubiquitination and degradation assays are provided in Supplementary Information.

Reporter gene assays

HEK 293T cells were grown in collagen-coated 24-well dishes and transfected using LipofectAMINE reagent (GIBCO BRL). After 36 h, cells were stimulated with EGF, lysed in 1 $\times$ lysis buffer (Pharmingen), and mixed with luciferase assay reagent (Pharmingen). The luciferase activity was measured by a luminometer (Wallace) and normalized against β -galactosidase activity for each point. EGF-induced increase in luciferase activity was quantified for every pair in the triplicate and was expressed as the average induction $\pm$ s.d.

Expression of transfected proteins was monitored for every experiment by immunoblotting. We repeated the experiments three times.

Received 8 November 2001; accepted 7 January 2002.

- Thien, C. B. & Langdon, W. Y. Cbl: many adaptations to regulate protein tyrosine kinases. *Nature Rev. Mol. Cell Biol.* **2**, 294–307 (2001).
- Joazeiro, C. A. et al. The tyrosine kinase negative regulator c-Cbl as a RING-type, E2-dependent ubiquitin-protein ligase. *Science* **286**, 309–312 (1999).
- Levkowitz, G. et al. Ubiquitin ligase activity and tyrosine phosphorylation underlie suppression of growth factor signaling by c-Cbl/Sli-1. *Mol. Cell* **4**, 1029–1040 (1999).
- Waterman, H. & Yarden, Y. Molecular mechanisms underlying endocytosis and sorting of ErbB receptor tyrosine kinases. *FEBS Lett.* **490**, 142–152 (2001).
- Di Fiore, P. P. & De Camilli, P. Endocytosis and signaling. An inseparable partnership. *Cell* **106**, 1–4 (2001).
- Take, H. et al. Cloning and characterization of a novel adaptor protein, CIN85, that interacts with c-Cbl. *Biochem. Biophys. Res. Commun.* **268**, 321–328 (2000).
- Schmidt, A. et al. Endophilin I mediates synaptic vesicle formation by transfer of arachidonate to lysophosphatidic acid. *Nature* **401**, 133–141 (1999).
- Ringstad, N. et al. Endophilin/SH3p4 is required for the transition from early to late stages in clathrin-mediated synaptic vesicle endocytosis. *Neuron* **24**, 143–154 (1999).
- Simpson, F. et al. SH3-domain-containing proteins function at distinct steps in clathrin-coated vesicle formation. *Nature Cell Biol.* **1**, 119–124 (1999).
- Gad, H. et al. Fission and uncoating of synaptic clathrin-coated vesicles are perturbed by disruption of interactions with the SH3 domain of endophilin. *Neuron* **27**, 301–312 (2000).
- Yoon, C. H., Chang, C., Hopper, N. A., Lesa, G. M. & Sternberg, P. W. Requirements of multiple domains of SLI-1, a *Caenorhabditis elegans* homologue of c-Cbl, and an inhibitory tyrosine in LET-23 in regulating vulval differentiation. *Mol. Biol. Cell* **11**, 4019–4031 (2000).
- Gout, I. et al. Negative regulation of PI 3-kinase by Ruk, a novel adaptor protein. *EMBO J.* **19**, 4015–4025 (2000).
- Bogler, O. et al. SETA: a novel SH3 domain-containing adapter molecule associated with malignancy in astrocytes. *Neuro-oncology* **2**, 6–15 (2000).
- Kirsch, K. H. et al. The adapter type protein CMS/CD2AP binds to the proto-oncogenic protein c-Cbl through a tyrosine phosphorylation-regulated Src homology 3 domain interaction. *J. Biol. Chem.* **276**, 4957–4963 (2001).
- Watanabe, S. et al. Characterization of the CIN85 adaptor protein and identification of components involved in CIN85 complexes. *Biochem. Biophys. Res. Commun.* **278**, 167–174 (2000).
- Borinstein, S. C. et al. SETA is a multifunctional adapter protein with three SH3 domains that binds Grb2, Cbl, and the novel SB1 proteins. *Cell Signal* **12**, 769–779 (2000).
- Ringstad, N., Nemoto, Y. & De Camilli, P. The SH3p4/Shp8/SH3p13 protein family: binding partners for synaptojanin and dynamin via a Grb2-like Src homology 3 domain. *Proc. Natl Acad. Sci. USA* **94**, 8569–8574 (1997).
- Micheva, K. D., Ramjaun, A. R., Kay, B. K. & McPherson, P. S. SH3 domain-dependent interactions of endophilin with amphiphysin. *FEBS Lett.* **414**, 308–312 (1997).
- Farsad, K. et al. Generation of high curvature membranes mediated by direct endophilin bilayer interactions. *J. Cell Biol.* **155**, 193–200 (2001).
- de Melker, A. A., van Der Horst, G., Calafat, J., Jansen, H. & Borst, J. c-Cbl ubiquitinates the EGF receptor at the plasma membrane and remains receptor associated throughout the endocytic route. *J. Cell Sci.* **114**, 2167–2178 (2001).
- Levkowitz, G. et al. c-Cbl/Sli-1 regulates endocytic sorting and ubiquitination of the epidermal growth factor receptor. *Genes Dev.* **12**, 3663–3674 (1998).
- Petrelli, A. et al. The endophilin–CIN85–Cbl complex mediates ligand-dependent downregulation of c-Met. *Nature* **416**, 187–190 (2001).
- Galisteo, M. L., Dikic, I., Batzer, A. G., Langdon, W. Y. & Schlessinger, J. Tyrosine phosphorylation of the c-cbl proto-oncogene protein product and association with epidermal growth factor (EGF) receptor upon EGF stimulation. *J. Biol. Chem.* **270**, 20242–20245 (1995).
- Scaife, R. M. & Langdon, W. Y. c-Cbl localizes to actin lamellae and regulates lamellipodia formation and cell morphology. *J. Cell Sci.* **113**(2), 215–226 (2000).
- Dikic, I., Tokiwa, G., Lev, S., Courtneidge, S. A. & Schlessinger, J. A role for Pyk2 and Src in linking G-protein-coupled receptors with MAP kinase activation. *Nature* **383**, 547–550 (1996).
- Ivankovic-Dikic, I., Gronroos, E., Blaukat, A., Barth, B. U. & Dikic, I. Pyk2 and FAK regulate neurite outgrowth induced by growth factors and integrins. *Nature Cell Biol.* **2**, 574–581 (2000).
- Sorkin, A., Mazzotti, M., Sorkina, T., Scotto, L. & Beguinot, L. Epidermal growth factor receptor interaction with clathrin adaptors is mediated by the Tyr974-containing internalization motif. *J. Biol. Chem.* **271**, 13377–13384 (1996).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank S. Giordano and Y. Yarden for communication of unpublished results, and L. Claesson-Welsh and C.-H. Heldin for critical reading of the manuscript. We also thank other members of the Dikic laboratory for valuable comments and help with various reagents. This work was supported by a Marie Curie Fellowship of the European Community programme to P.S., and the Boehringer Ingelheim Fonds and the Swedish Strategic Funds to I.D.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.D. (e-mail: Ivan.Dikic@licr.uu.se).

The endophilin–CIN85–Cbl complex mediates ligand-dependent downregulation of c-Met

Annalisa Petrelli[†], Giorgio F. Gilestro[†], Stefania Lanzardo^{*}, Paolo M. Comoglio[†], Nicola Migone^{*‡} & Silvia Giordano^{†‡}

^{*} CNR-CIOS and Department of Genetics, Biology and Biochemistry, University of Torino, 10126 Torino, Italy

[†] Division of Molecular Oncology, Institute for Cancer Research and Treatment (IRCC), University of Torino Medical School, I-10060 Candiolo, Torino, Italy

[‡] These authors contributed equally to this work

Ligand-dependent downregulation of tyrosine kinase receptors is a critical step for modulating their activity. Upon ligand binding, hepatocyte growth factor (HGF) receptor (Met) is polyubiquitinated¹ and degraded²; however, the mechanisms underlying HGF receptor endocytosis are not yet known. Here we demonstrate that a complex involving endophilins, CIN85 and Cbl controls this process. Endophilins³ are regulatory components of clathrin-coated vesicle formation. Through their acyl-transferase activity they are thought to modify the membrane phospholipids and induce negative curvature and invagination of the plasma membrane during the early steps of endocytosis⁴. Furthermore, by means of their Src-homology 3 domains, endophilins are able to bind CIN85, a recently identified protein that interacts with the Cbl proto-oncogene⁵. Cbl, in turn, binds and ubiquitinates activated HGF receptor, and by recruiting the endophilin–CIN85 complex, it regulates receptor internalization. Inhibition of complex formation is sufficient to block HGF receptor internalization and to enhance HGF-induced signal transduction and biological responses. These data provide further evidence of a relationship between receptor-mediated signalling and endocytosis, and disclose a novel functional role for Cbl in HGF receptor signalling.

Cell stimulation with growth factors usually leads to increased receptor tyrosine kinase activity, intracellular substrate recruitment, receptor internalization by accelerated endocytosis through clathrin-coated pits, and finally, degradation^{6–8}. Several data sets suggest that ligand-dependent internalization may be a principal process regulating the duration and propagation of the signal initiated by tyrosine kinase receptors, thereby preventing overstimulation that could potentially lead to cellular transformation.

To clarify the molecular mechanisms connecting receptor-mediated signalling to endocytosis by means of clathrin-coated pits, we searched for proteins interacting with endophilins, by using the yeast two-hybrid system. A bait construct, encoding the full-length human endophilin A3, fused in-frame with the Gal4 DNA-binding domain, was used to screen a human brain complementary DNA library. Approximately 500,000 transformants were analysed, leading to isolation of 27 positive colonies. Sequence analysis of the prey plasmids revealed that five of them encoded for dynamins (well known interactors of endophilins)⁹, whereas three clones of different length represented overlapping cDNAs for CIN85 (Fig. 1a). CIN85 is a member of a newly discovered subfamily of adaptor molecules including CMS¹⁰ and CD2AP¹¹, and was recently shown to bind the Cbl proto-oncogene product⁵. CIN85 contains three amino-terminal Src-homology 3 (SH3) domains, followed by a proline-rich domain containing two sequences (PKKPPPP and PKKPRPP) that are optimal consensus sites for endophilin SH3 binding¹² and a carboxy-terminal coiled-coil domain, essential for dimerization¹³. As shown in Fig. 1b, pull-down experiments confirmed that a recombinant glutathione S-transferase (GST) fusion protein encoding full-length endophilin A3 was able to interact with

CIN85, and that the SH3 domains of the known endophilins are necessary and sufficient for this binding. Co-immunoprecipitation experiments demonstrated that endophilin A3 and CIN85 also associate in mammalian cells (Supplementary Information Fig. 1). As endophilin A3 binds CIN85, which in turn interacts with Cbl⁵, we investigated whether these three proteins can form a ternary complex in cells. As shown in Fig. 1c, in transfected cells endophilin A3, CIN85 and Cbl associate in a complex; moreover, a dominant interfering form of CIN85, which is unable to interact with Cbl (CIN85-PCC, containing only the proline-rich and the coiled-coil domains), prevents complex formation.

Cbl binds to several activated tyrosine kinase receptors, acting either as a transducer¹⁴ or as a ubiquitin ligase, and thus has a role in receptor downregulation¹⁵. In particular, the role of Cbl has been studied in signal transduction activated by HGF receptor¹⁶, the heterodimeric tyrosine kinase with a relative molecular mass of 190,000 (190K) that is encoded by the Met proto-oncogene¹⁷. After HGF binding, signalling is elicited through receptor interactions with the major substrate Gab1 (ref. 18) and several SH2-containing proteins¹⁹, one of which is the Cbl adaptor protein, known to bind to the Met tail¹⁶ and to have a positive role in mitogen-activated protein (MAP) kinase activation²⁰. After ligand stimulation, Met is rapidly polyubiquitinated¹ and degraded²; however, the molecular mechanisms involved in this process are unknown. We therefore investigated whether Cbl has a role in HGF receptor ubiquitination.

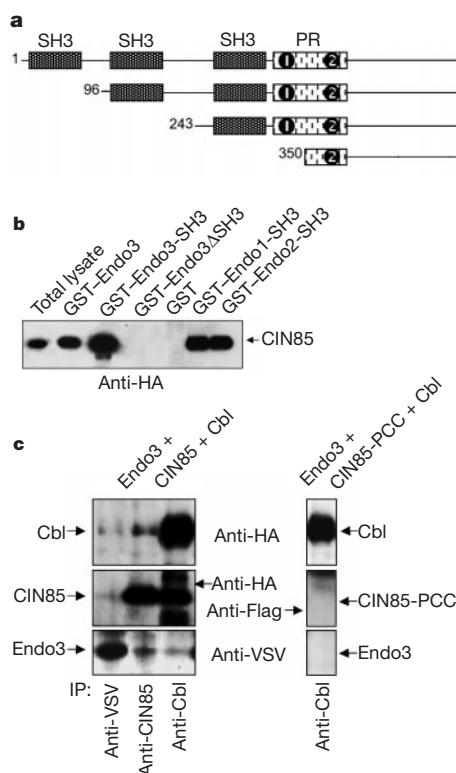


Figure 1 Endophilins interact with CIN85 and Cbl. **a**, Schematic representation of full-length CIN85 and the three clones obtained from the yeast two-hybrid screening. PR, proline-rich; CC, coiled-coil. Numbers inside the proline-rich domain indicate putative consensus sequences for endophilin SH3 binding. **b**, Pull-down experiments performed on lysates of cells transfected with HA-CIN85. CIN85 interacts with full-length endophilin A3 (GST-Endo3), the SH3 domains of endophilin A1, A2 and A3, but not with endophilin A3 lacking the SH3 domain (GST-Endo3ΔSH3). **c**, Endophilin A3 associates in a complex with CIN85 and Cbl. Lysates of HEK 293T cells co-transfected with VSV-endophilin A3, CIN85 (HA-wild type (WT) or Flag-CIN85-PCC) and HA-Cbl were immunoprecipitated with the indicated antibodies. Expression of a CIN85 mutant unable to bind Cbl (CIN85-PCC) impairs complex formation.

As shown in Fig. 2a, Cbl binds to the receptor (right panel) and induces its ubiquitination (left panel). On the contrary, a Cbl mutant devoid of ubiquitin ligase activity (Cbl-70Z) interacts with Met, but is unable to induce its ubiquitination (left panel). Notably, this mutant does not act in a dominant interfering manner with respect to preventing Met downregulation (Supplementary Information Fig. 2B). The interaction between Cbl and Met requires the presence of Y1356 in the receptor docking site¹⁶. We show here that a Cbl mutant lacking the SH2 domain is still able to associate with Met, although to a lesser extent, and to ubiquitinate the receptor (Supplementary Information Fig. 2A). It is possible that the interaction between this Cbl mutant and Met might occur indirectly by means of a bridge made by the SH3 domain of the adaptor protein Grb2 and the Cbl proline-rich region. This alternative modality of interaction has recently been observed also for Cbl and the EGF receptor (Y. Yarden, personal communication).

We then investigated whether the association of endophilin, CIN85 and Cbl was modulated by HGF. Indeed, under physiological conditions, Cbl binds to ligand-activated HGF receptor, becomes tyrosine phosphorylated, and recruits the CIN85-endophilin complex (Fig. 2b). Furthermore, endophilins are constitutively associated to HGFR in a cell line (GTL16) where Met is basally activated¹⁷ (Supplementary Information Fig. 3).

Endophilins are known to be regulatory elements of the machinery controlling endocytosis through clathrin-coated pits^{3,21-23}. To verify whether the Cbl-CIN85-endophilin complex recruited to Met has a role in receptor internalization, we impaired complex formation by using dominant interfering forms of each component of the complex. As shown by immunofluorescence analysis (Fig. 3a, b), the HGF-HGFR complex is not internalized in cells expressing mutants of Cbl (Cbl-480, a mutant that is unable to interact with CIN85; I. Dikic, personal communication), CIN85 (CIN85-PCC) and endophilin (Endo-SH3, a mutant still able to bind CIN85 but devoid of enzymatic activity). Indeed, in these cells, ligand-induced

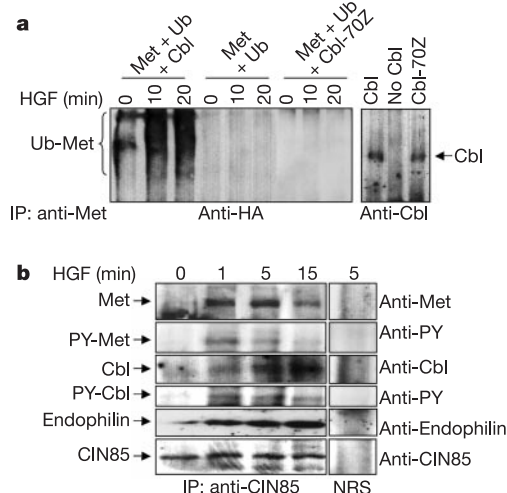


Figure 2 Cbl ubiquitinates Met and recruits the CIN85-endophilin complex. **a**, Activated Met interacts with Cbl and undergoes ubiquitination. HEK 293T cells were co-transfected with Met and HA-tagged ubiquitin (Ub) either in the absence or in the presence of wild-type Cbl (Cbl) or of a Cbl mutant devoid of ubiquitin ligase activity (Cbl-70Z). Cells were stimulated with HGF and immunoprecipitated with anti-Met antibodies. Activated Met associates with Cbl (right panel) and becomes polyubiquitinated (left panel). Cbl-70Z interacts with Met but it is not able to induce Met ubiquitination. **b**, Ligand-activated Met recruits the Cbl-CIN85-endophilin complex. HeLa cells were stimulated with HGF. Cell lysates were immunoprecipitated with either anti-CIN85 antibodies or pre-immune serum (NRS). Upon HGF-binding, Met becomes tyrosine phosphorylated, and binds to and phosphorylates Cbl, inducing the association with CIN85 and endophilin. PY-Met, tyrosine phosphorylated Met; PY-Cbl, tyrosine phosphorylated Cbl.

HGF receptor downregulation is strongly impaired—the amount of receptor is increased even in basal conditions and remains almost unaltered on HGF stimulation (Fig. 3c). Moreover, HGFR tyrosine phosphorylation is prolonged. Of note, in the presence of dominant interfering forms of either CIN85 or endophilin, the interaction between Cbl and Met is not impaired, Cbl becomes tyrosine phosphorylated (Supplementary Information Fig. 4), and Cbl-induced HGFR ubiquitination is unaffected (Supplementary Information Fig. 2A and data not shown).

To assess the signalling capacity of HGFR not undergoing downregulation, we evaluated its ability to recruit signal transducers in cells expressing the interfering form of endophilin. We found that when the receptor is not internalized, it remains tyrosine phos-

phorylated for a longer time (Fig. 3c), increasing binding and phosphorylation of its major substrate Gab1 (Fig. 4a). This increase of substrate recruitment and phosphorylation leads to enhancement of HGF-induced motility (Fig. 4b). We also observed a basal increase of motogenic properties, probably due to the fact that blocking of ligand-induced endocytosis may affect several receptors²⁴, and thus enhance cell responsivity to serum chemotactic molecules.

To explain our observations we propose the following model. Upon ligand activation, HGF receptor becomes tyrosine phosphorylated, binds and phosphorylates Cbl, which in turn targets the receptor to clathrin-coated pits by recruiting the CIN85–endophilin complex. This complex drives plasma membrane invagination and vesicle formation, thus resulting in negative modulation of signal transduction and biological responses. Our results also show that Cbl has a dual role in receptor signalling: it acts as a positive transducer, by triggering signals from the activated receptor, and as a negative modulator, by ubiquitinating the receptor and activating the endocytic machinery through recruitment of CIN85 and endophilin. Moreover, this role of Cbl in induction of endocytosis can be functionally separated from its E3 ligase activity. As a similar role for the Cbl–CIN85–endophilin complex has been demonstrated for downregulation of EGF receptors²⁴, the mechanism we describe seems to represent a general way to downregulate activated tyrosine kinase receptors. □

Methods

Plasmid construction and yeast two-hybrid screening

For details of plasmid construction and yeast two-hybrid screening see Supplementary Information.

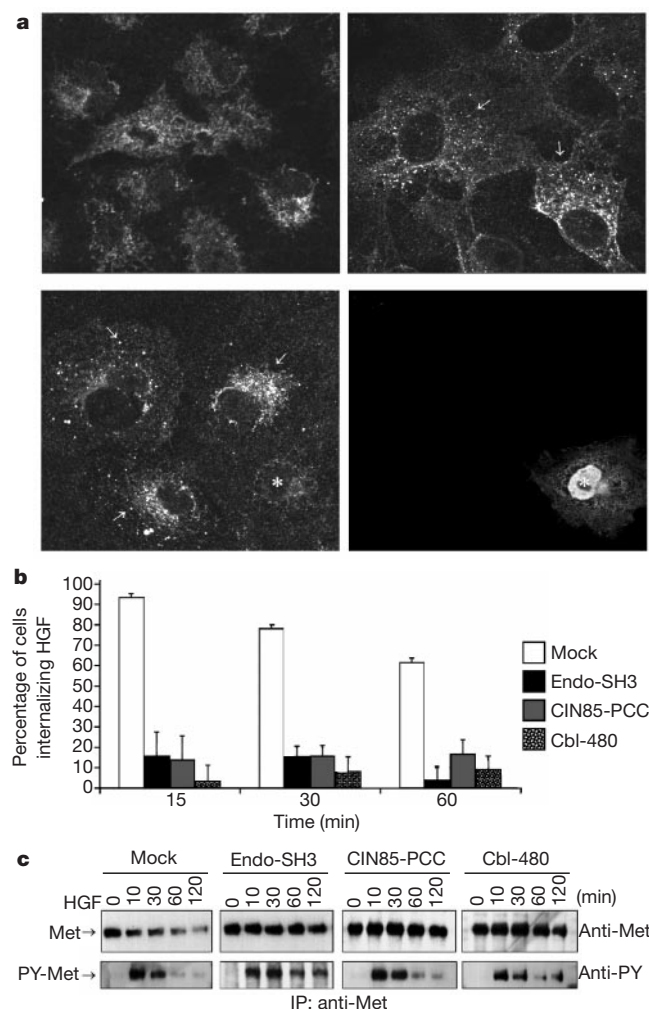


Figure 3 Expression of dominant interfering mutants of Cbl, CIN85 and endophilin impairs HGF receptor internalization and downregulation. **a**, Confocal microscopy images of COS7 cells transfected with Mock (top panels) or endophilin A3 SH3 (bottom panels), treated (top right and bottom panels) or not (top left panel) with HGF (30 min). The top and bottom left panels show staining with anti-HGF antibodies; the bottom right panel (same field as in bottom left, asterisk) is stained to determine endophilin SH3 expression. Ligand internalization (fluorescent dots, some of which are indicated by arrows) is abrogated in cells expressing endophilin SH3 domain. Magnification is at $\times 100$. **b**, Quantitative analysis of immunofluorescence experiments. COS7 cells were transfected with Cbl, CIN85 or endophilin mutants. Immunofluorescence was performed as in **a**. The percentage of cells containing fluorescent dots was evaluated. Experimental points were carried out in triplicate. **c**, HEK 293T cells expressing the same mutants were stimulated with HGF. The same amount of total proteins was immunoprecipitated with anti-Met antibodies. In these cells, HGFR downregulation is impaired, and receptor tyrosine phosphorylation is prolonged.

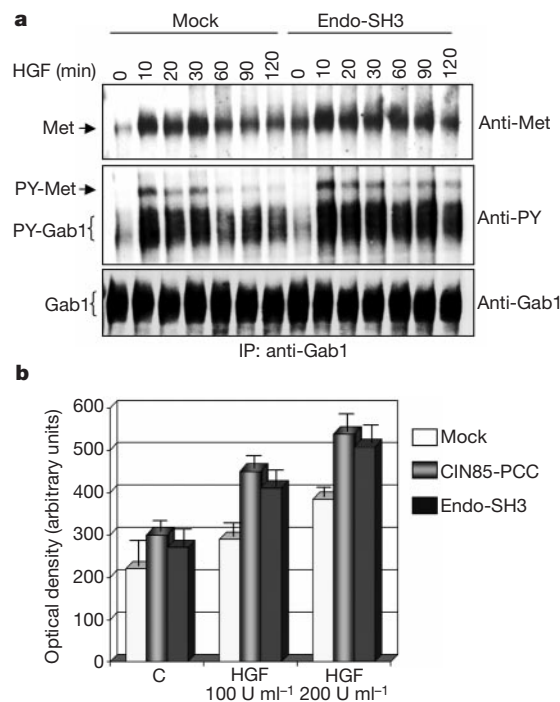


Figure 4 Impairment of Met downregulation enhances signal transduction and biological responses. **a**, A dominant interfering form of endophilin A3 enhances HGF-induced signal transduction. 293T cells transfected with Mock or Endo-SH3 were stimulated with HGF. Equal amounts of proteins were immunoprecipitated with anti-Gab1 antibodies. Endo-SH3-expressing cells show an increased association of Gab1 with Met and prolonged tyrosine phosphorylation. **b**, Dominant interfering forms of CIN85 or endophilin A3 increase cell motility. HEK 293T cells were transfected with Mock, CIN85 or endophilin mutants, and chemotactic ability of HGF was evaluated in a modified Boyden chamber assay. For experimental details see Methods.

Transfection, immunoprecipitation and western blotting

Human embryonic kidney (HEK) 293T and COS7 cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum. Cells were co-transfected with 3 µg of each plasmid and cultured for 48 h. For *in vivo* stimulation experiments, cells were grown for 48 h in spent medium and then stimulated with HGF (400 U ml⁻¹) at 37 °C, for the indicated times. Ubiquitination experiments were performed as described¹⁵.

For immunoprecipitations, cells were lysed with EB buffer (20 mM Tris-HCl, pH 7.4, 5 mM EDTA, 150 mM NaCl, 10% glycerol, 1% Triton X-100) in the presence of protease inhibitors and 1 mM Na-orthovanadate. After immunoprecipitation with the appropriate antibodies, high stringency washes were performed (EB plus 1 M LiCl).

Western blots were performed according to standard methods. In western blot and immunoprecipitations the following antibodies were used: PY20 monoclonal antibody (Transduction Laboratories), anti-Met antibodies (described in ref. 25), anti vesicular stomatitis virus (VSV)-G (Sigma), anti-Cbl and anti-haemagglutinin (HA) (Santa Cruz Biotechnology), anti-Gab1 (Upstate Biotechnology), and anti-Flag (Sigma). Monoclonal anti-endophilin A3 and polyclonal anti-endophilins antibodies (able to recognize endophilins A1, A2 and A3) were a gift of E. Lantelme and C. Giachino. Anti-CIN85 (CT) antibody and pre-immune serum were provided by I. Dikic. Final detection was done with ECL system (Amersham).

GST pull-down assay

The host strain *Escherichia coli* XL-1 was transformed with plasmids encoding GST fusion proteins, and protein expression was induced with isopropyl-β-D-thiogalactopyranoside. Bacterial cells were lysed in phosphate-buffered saline containing 1% Triton X-100 and protease inhibitors. Fusion proteins were captured on glutathione-Sepharose beads and incubated with crude lysates of transfected HEK 293T cells at 4 °C for 2 h. The beads were washed extensively in lysis buffer, and then bound proteins were eluted by boiling in SDS-polyacrylamide gel electrophoresis (PAGE) sample buffer, separated by SDS-PAGE, and processed for western blotting.

Motility assay

For motility assay, 10⁵ cells were seeded on the upper side of a Transwell chamber on a porous polycarbonate membrane (8.0-µm pore size); the lower chamber of the Transwell was filled with DMEM containing 5% FBS, either in the absence or in the presence of HGF (100 or 200 U ml⁻¹). After 24 h of incubation, cells attached to the upper side of the filter were mechanically removed. Cells that migrated to the lower side were fixed and stained with crystal violet. The stained cells were solubilized in 10% acetic acid, absorbance at 560 nm was measured in a micro-plate reader.

Immunofluorescence

COS7 cells were transfected using Lipofectamine (Life Technologies) and then seeded on polyisiline-coated coverslips. Cells were treated with HGF (400 U ml⁻¹) for different times, then fixed in 4% paraformaldehyde, permeabilized in 0.1% Triton X-100 and finally co-stained using biotin-conjugated anti-HGF (R&D), anti-VSV to detect endophilin A3 SH3 domain, anti-Flag to detect CIN85-PCC, and anti-HA to detect Cbl-480. Images were taken using a Bio-Rad Confocal Microscopy System.

Received 12 November 2001; accepted 7 January 2002.

- Jeffers, M., Taylor, G. A., Weidner, K. M., Omura, S. & Vande Woude, G. F. Degradation of the Met tyrosine kinase receptor by the ubiquitin-proteasome pathway. *Mol. Cell. Biol.* **17**, 799–808 (1997).
- Hammond, D. E., Urbé, S., Vande Woude, G. F. & Clague, M. J. Down-regulation of MET, the receptor for hepatocyte growth factor. *Oncogene* **20**, 2761–2770 (2001).
- Giachino, C. *et al.* A novel SH-3-containing human gene family preferentially expressed in the central nervous system. *Genomics* **41**, 427–434 (1997).
- Schmidt, A. *et al.* Endophilin I mediates synaptic vesicle formation by transfer of arachidonate to lysophosphatidic acid. *Nature* **401**, 133–141 (1999).
- Take, H. *et al.* Cloning and characterization of a novel adaptor protein, CIN85, that interacts with c-Cbl. *Biochem. Biophys. Res. Commun.* **268**, 321–328 (2000).
- Hunter, T. Signaling—2000 and beyond. *Cell* **100**, 113–127 (2000).
- Waterman, H. & Yarden, Y. Molecular mechanisms underlying endocytosis and sorting of ErbB receptor tyrosine kinases. *FEBS Lett.* **490**, 142–152 (2001).
- Di Fiore, P. P. & De Camilli, P. Endocytosis and signaling: an inseparable partnership. *Cell* **106**, 1–4 (2001).
- Ringstad, N., Nemoto, Y. & De Camilli, P. The SH3p4/SH3p8/SH3p13 protein family: binding partners for synaptotagmin and dynamin via a Grb2-like Src homology 3 domain. *Proc. Natl Acad. Sci. USA* **94**, 8569–8574 (1997).
- Kirsch, K. H., Georgescu, M. M., Ishimura, S. & Hanafusa, H. CMS: an adapter molecule involved in cytoskeletal rearrangements. *Proc. Natl Acad. Sci. USA* **99**, 6211–6216 (1999).
- Dustin, M. L. *et al.* A novel adaptor protein orchestrates receptor patterning and cytoskeletal polarity in T-cell contacts. *Cell* **94**, 667–677 (1998).
- Cestra, G. *et al.* The SH3 domains of Endophilin and Amphiphysin bind to the proline rich region of Synaptotagmin 1 at distinct sites that display an unconventional binding specificity. *J. Biol. Chem.* **274**, 32001–32007 (1999).
- Watanabe, S. *et al.* Characterization of the CIN85 adaptor protein and identification of components involved in CIN85 complexes. *Biochem. Biophys. Res. Commun.* **278**, 167–174 (2000).
- Thien, C. B. F. & Langdon, W. Y. Cbl: many adaptations to regulate protein tyrosine kinases. *Nature Rev. Mol. Cell Biol.* **2**, 294–307 (2001).
- Levkowitz, G. *et al.* Ubiquitin ligase activity and tyrosine phosphorylation underlie suppression of growth factor signaling by c-Cbl/Sli-1. *Mol. Cell* **4**, 1029–1040 (1999).
- Fixman, E. D. *et al.* Efficient cellular transformation by the Met oncoprotein requires a functional

- Grb2 binding site and correlates with phosphorylation of the Grb2-associated proteins, Cbl and Gab1. *J. Biol. Chem.* **272**, 20167–20172 (1997).
- Giordano, S., Ponzetto, C., Di Renzo, M. F., Cooper, C. S. & Comoglio, P. M. Tyrosine kinase receptor indistinguishable from the c-met protein. *Nature* **339**, 155–156 (1989).
- Weidner, K. M. *et al.* Interaction between Gab1 and the c-Met receptor tyrosine kinase is responsible for epithelial morphogenesis. *Nature* **384**, 173–176 (1996).
- Ponzetto, C. *et al.* A multifunctional docking site mediates signaling and transformation by the hepatocyte growth factor/scatter factor receptor family. *Cell* **77**, 261–271 (1994).
- Garcia-Guzman, M., Larse, E. & Vuori, K. The proto-oncogene c-Cbl is a positive regulator of Met-induced MAP kinase activation: a role for the adaptor protein Crk. *Oncogene* **19**, 4058–4065 (2000).
- Brodin, L., Low, P. & Shupliakov, O. Sequential steps in clathrin-mediated synaptic vesicle endocytosis. *Curr. Opin. Neurobiol.* **10**, 312–320 (2000).
- Huttner, W. B. & Schmidt, A. Lipids, lipid modification and lipid-protein interaction in membrane budding and fission—insights from the roles of endophilin A1 and synaptophysin in synaptic vesicle endocytosis. *Curr. Opin. Neurobiol.* **10**, 543–551 (2000).
- Hill, E., van Der Kaay, J., Downes, C. P. & Smythe, E. The role of dynamin and its binding partners in coated pit invagination and scission. *J. Cell Biol.* **152**, 309–323 (2001).
- Soubeyran, P., Kowanetz, K., Szymkiewicz, L., Langdon, W. Y. & Dikic, I. Cbl–CIN85–endophilin complex mediates ligand-induced downregulation of EGF receptors. *Nature* **416**, 183–187 (2001).
- Prat, M., Crepaldi, T., Pennacchietti, S., Bussolino, F. & Comoglio, P. M. Agonistic monoclonal antibodies against the Met receptor dissect the biological responses to HGF. *J. Cell Sci.* **111**, 237–247 (1998).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank E. Lantelme, C. Giachino and L. Lanzetti for production of anti-endophilin antibodies and for GST–Endo3, GST–Endo1-2-SH3 constructs; L. Tamagnone for ΔPlexin A3 cDNA; P. De Camilli for dynamin cDNA; Y. Yarden for Cbl and ubiquitin plasmids; I. Dikic for CIN85 reagents and Cbl-ΔSH2 plasmid; and W. Y. Langdon for Cbl-70Z and Cbl-480 constructs. We also thank Y. Yarden, L. Tamagnone and our colleagues for discussions; A. Cignetto for secretarial assistance; and E. Wright for editing the manuscript. The technical assistance of L. Palmas, R. Albano and E. Clemente is acknowledged. This work was partly supported by Associazione E. & E. Rulfo per la Genetica Medica to N.M., by Italian Association for Cancer Research to P.M.C. and by MURST COFIN to S.G.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.G. (e-mail: sgiordano@ircc.unito.it).

Involvement of receptor-interacting protein 2 in innate and adaptive immune responses

Arnold I. Chin^{†‡}, Paul W. Dempsey^{†‡}, Kevin Bruhn[‡], Jeff F. Miller^{†‡}, Yang Xu[§] & Genhong Cheng^{*‡}

* Molecular Biology Institute, University of California, Los Angeles, California 90095, USA

‡ Department of Microbiology, Immunology, and Molecular Genetics, Jonsson Comprehensive Cancer Center, University of California, Los Angeles, California 90095, USA

§ Section of Molecular Biology, Division of Biology, University of California, San Diego, California 92093, USA

† These authors contributed equally to this work

Host defences to microorganisms rely on a coordinated interplay between the innate and adaptive responses of immunity¹. Infection with intracellular bacteria triggers an immediate innate response requiring macrophages, neutrophils and natural killer cells, whereas subsequent activation of an adaptive response through development of T-helper subtype 1 cells (T_H1) proceeds during persistent infection¹. To understand the physiological

role of receptor-interacting protein 2 (Rip2), also known as RICK and CARDIAK, we generated mice with a targeted disruption of the gene coding for Rip2. Here we show that Rip2-deficient mice exhibit a profoundly decreased ability to defend against infection by the intracellular pathogen *Listeria monocytogenes*. Rip2-deficient macrophages infected with *L. monocytogenes* or treated with lipopolysaccharide (LPS) have decreased activation of NF- κ B, whereas dominant negative Rip2 inhibited NF- κ B activation mediated by Toll-like receptor 4 and Nod1. *In vivo*, Rip2-deficient mice were resistant to the lethal effects of LPS-induced endotoxic shock. Furthermore, Rip2 deficiency results in impaired interferon- γ production in both T_H1 and natural killer cells, attributed in part to defective interleukin-12-induced Stat4 activation. Our data reflect requirements for Rip2 in multiple pathways regulating immune and inflammatory responses.

Rip2 contains an amino-terminal kinase domain and a carboxy-terminal caspase activation and recruitment domain (CARD)²⁻⁴. We disrupted the *Rip2* gene in mice by replacing the first exon of Rip2 with a neomycin-resistance gene (Fig. 1a). Targeted disruption was determined by Southern analysis (Fig. 1b), and loss of Rip2 expression was demonstrated by western analysis (Fig. 1c). *Rip2*^{-/-} mice were viable and fertile, and have survived up to 24 months with no gross differences compared to *Rip2*^{+/+} littermates. Offspring genotypes from heterozygous breeding approached mendelian frequency. Development and ratios of all haematopoietic lineages tested appeared equivalent in *Rip2*^{+/+} and *Rip2*^{-/-} mice by analysis of typical surface markers. Overexpressed Rip2 has been shown to promote apoptosis^{2,3}; however, sensitivity to apoptosis induced by a variety of stimuli was similar in *Rip2*^{+/+} and *Rip2*^{-/-} embryonic fibroblast cells and thymocytes (data not shown).

As Rip2 interacts with CD40, a co-stimulatory receptor critical for T-cell-dependent humoral responses², we assessed the role of Rip2 in the humoral immune response. On immunization of mice with the T-cell-dependent antigen chicken egg ovalbumin (OVA), we observed that titres of OVA-specific immunoglobulin- γ (IgG)-2a were up to 20-fold lower in *Rip2*^{-/-} mice compared with

Rip2^{+/+} mice (Fig. 2a). In contrast, OVA-specific IgG1 titres were similar in *Rip2*^{+/+} and *Rip2*^{-/-} mice (Fig. 2b). No significant differences in B-cell proliferation were found between *Rip2*^{+/+} and *Rip2*^{-/-} B cells stimulated with anti-CD40 and anti-IgM (Fig. 2c, d). On the other hand, decreased proliferation was observed in *Rip2*^{-/-} CD4⁺ T cells stimulated with anti-CD3 compared with *Rip2*^{+/+} CD4⁺ T cells (Fig. 2e). As a role for Rip2 in T cells was unexpected, we also examined *Rip2* gene expression, and detected increasing levels of Rip2 messenger RNA during activation of T cells (Fig. 2f).

To explore the ability of Rip2 to regulate T-cell responses, we immunized *Rip2*^{+/+} and *Rip2*^{-/-} mice with OVA in complete Freund's adjuvant (CFA)⁵. Consistent with the decreased immunoglobulin isotype switching to IgG2a, lower amounts of IFN- γ were produced by *Rip2*^{-/-} draining lymph node cells restimulated with antigen (Fig. 3a; see also ref. 6). These data suggest impaired T_H1 activation in *Rip2*^{-/-} mice, a function critical for the generation of cell-mediated and inflammatory responses to pathogen challenge^{7,8}. To further investigate T-helper cell development, we cultured *Rip2*^{+/+} and *Rip2*^{-/-} CD4⁺ T cells under polarizing conditions to promote either T_H1 or T_H2 development. When restimulated with anti-CD3, T_H1 cells of *Rip2*^{-/-} mice produced twofold less IFN- γ than T_H1 cells of *Rip2*^{+/+} mice (Fig. 3b). In contrast, similar levels of interleukin (IL)-4 were secreted in *Rip2*^{-/-} and *Rip2*^{+/+} T_H2 -polarized cells, suggesting that *Rip2*^{-/-} T cells have an impaired ability to differentiate into T_H1 cells, but a normal capacity to differentiate into T_H2 cells (Fig. 3c).

To specifically study T_H1 cells, we generated T_H1 clones. In CD4⁺ T cells, IFN- γ production can be induced either through T-cell

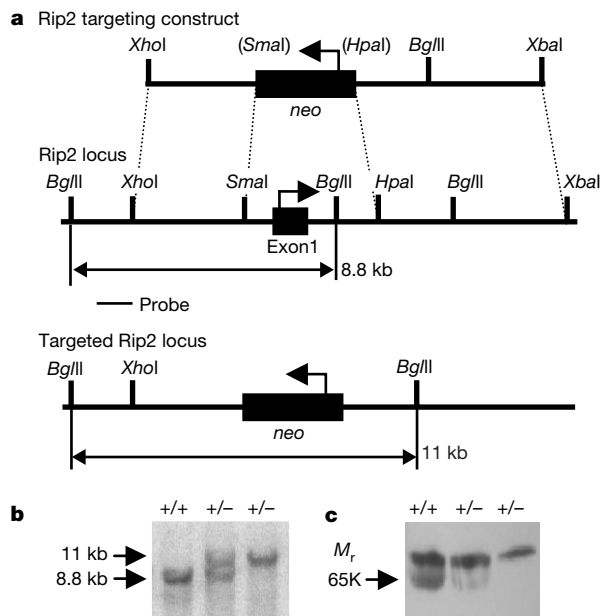


Figure 1 Generation of Rip2-deficient mice. **a**, Schematic representation of the gene-targeting construct (top), the genomic Rip2 locus (middle), and the targeted Rip2 locus (bottom). The probe for Southern analysis is indicated. **b**, Genomic tail DNA was digested with *Bgl*II, and blots were hybridized with the probe indicated in **a**. **c**, Extracts from splenocytes of *Rip2*^{+/+}, *Rip2*^{+/-}, and *Rip2*^{-/-} mice were examined by western analysis using a Rip2-specific antibody. *M_r*, relative molecular mass.

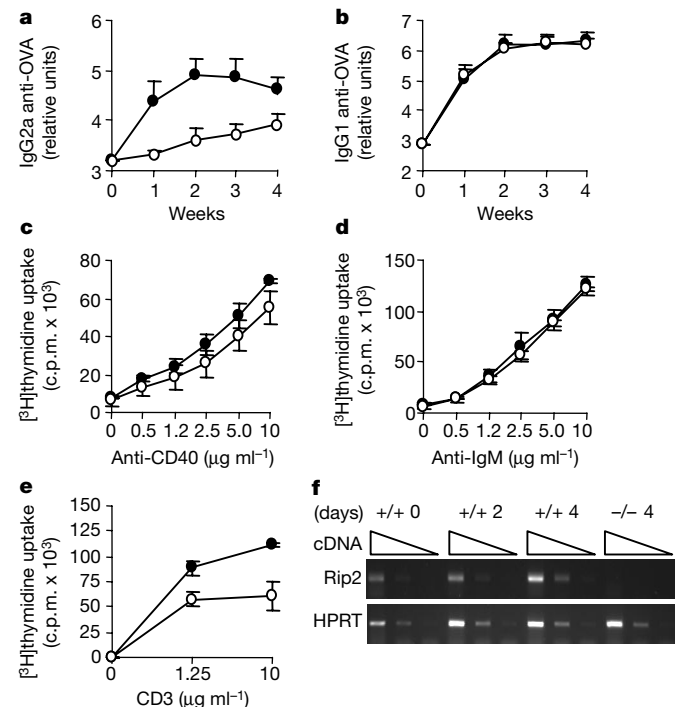


Figure 2 Impaired isotype switching and T-cell proliferation in *Rip2*^{-/-} cells. **a, b**, Sera from *Rip2*^{+/+} (filled circles) and *Rip2*^{-/-} (open circles) mice immunized with OVA (100 µg) adsorbed on alum were used to measure levels of IgG2a (**a**) or IgG1 (**b**) using isotype-specific antibodies by ELISA. *y* axis is on a logarithmic scale. **c, d**, Purified *Rip2*^{+/+} (filled circles) or *Rip2*^{-/-} (open circles) splenic B cells were stimulated with anti-CD40 (**c**) or anti-IgM (**d**) for 48 h, and assessed for proliferation. **e**, Purified *Rip2*^{+/+} (filled circles) and *Rip2*^{-/-} (open circles) CD4⁺ T cells were stimulated with anti-CD3 for 48 h. Proliferation was assessed by addition of [³H]thymidine. **f**, Serial dilutions of template cDNA (1:1, 1:10, 1:100) were used for semi-quantitative RT-PCR measuring Rip2 expression during T-cell differentiation after 0, 2, and 4 days for *Rip2*^{+/+}, and 4 days for *Rip2*^{-/-} CD4⁺ T cells.

antigen receptor (TCR) or IL-12/IL-18 activation⁹. To delineate the specific pathway involving Rip2 in IFN- γ regulation, we stimulated T_H1 clones with IL-12, IL-18, or both, and measured IFN- γ production. Whereas IL-12 or IL-18 alone did not activate IFN- γ secretion, *Rip2*^{-/-} T_H1 clones produced 76% less IFN- γ on average after synergistic activation with the combination of IL-12 and IL-18, compared with *Rip2*^{+/+} clones. After stimulation with anti-CD3 alone, or with IL-12, IL-18, or IL-12 plus IL-18, in conjunction with TCR activation, *Rip2*^{-/-} clones produced 34%, 43%, 45% or 36% less IFN- γ on average, respectively, than *Rip2*^{+/+} cells (Fig. 3d). Statistically significant differences in IL-12-mediated IFN- γ production led us to examine further the role of Rip2 in IL-12 signalling.

IL-12 mediates its function largely through the activation of signal transducer and activator of transcription 4 (Stat4). Indeed, *Stat4*^{-/-} T_H1 cells are defective in IL-12-induced IFN- γ production^{10,11}. Similar expression levels of IL-12 receptor in *Rip2*^{+/+} and *Rip2*^{-/-} clones, as measured by western blot and polymerase chain reaction with reverse transcription (RT-PCR) (data not shown), implicated Rip2 as functioning downstream of the IL-12 receptor. To determine whether the loss of Rip2 affected Stat4 activation, we stimulated *Rip2*^{+/+} and *Rip2*^{-/-} T_H1 clones with IL-12, and measured activation of Stat4. *Rip2*^{-/-} T_H1 clones exhibited impaired Stat4 activation, as shown by decreased tyrosine phosphorylation and decreased DNA binding, whereas total

amounts of Stat4 remained constant (Fig. 3e). Thus, defective Stat4 activation may contribute towards impaired IFN- γ production in *Rip2*^{-/-} mice.

The inflammatory cytokine IL-12 bridges innate and adaptive immunity, in part, by induction of IFN- γ , a critical mediator of T_H1, cytotoxic T-lymphocyte (CTL), and natural killer (NK) cell responses¹². We next investigated whether impaired IL-12-mediated IFN- γ production in T_H1 cells of *Rip2*^{-/-} mice could be a generalized defect, and examined the ability of NK cells to respond to IL-12. Splenocytes from *Rip2*^{+/+} and *Rip2*^{-/-} mice were stimulated with IL-12, IL-18, or both IL-12 and IL-18. Whereas splenocytes of *Rip2*^{+/+} mice secreted IFN- γ after stimulation with IL-12 or IL-18, splenocytes of *Rip2*^{-/-} mice did not secrete IFN- γ at levels above background under similar conditions, demonstrating an impairment of IL-12 and IL-18 signalling in NK cells. Synergistic activation with IL-12 and IL-18 induced twofold less IFN- γ from *Rip2*^{-/-} cells (Fig. 3f). Impairments in IL-18-mediated signalling were observed in both T_H1 and NK cells. This finding indicates an additional role for Rip2 in the related Toll-like receptor (TLR)/IL-1/IL-18 superfamily.

IFN- γ is crucial in the defence against intracellular pathogens such as *L. monocytogenes*, as mice deficient in IFN- γ , IFN- γ R, or Stat4 have increased susceptibility to infection by *L. monocytogenes*^{11,13,14}. To demonstrate the *in vivo* role of Rip2 in immune responses, we used a 10% LD₅₀ dose (lethal to 50% of animals tested) of *L. monocytogenes* to intravenously infect *Rip2*^{+/+} and *Rip2*^{-/-} mice. Although the bacterial titres among *Rip2*^{+/+} and *Rip2*^{-/-} mice after 1.5 days were similar in both livers and spleens, after 5 days the numbers of *L. monocytogenes* recovered from the livers and spleens of *Rip2*^{-/-} mice were about 75- and 25-fold higher than those in *Rip2*^{+/+} mice, respectively (Fig. 4a, b). Indeed, all *Rip2*^{-/-} mice succumbed to infection within 8 days, whereas 83% of *Rip2*^{+/+} mice survived and completely cleared the bacteria by 10 days (Fig. 4c).

Interestingly, up to tenfold higher *L. monocytogenes* titres were found in *Rip2*^{-/-} mice 3 days after infection, earlier than the 4–5 days required for the development of an adequate adaptive response, suggesting an intrinsic defect in innate immunity as well. Thus, to examine the intrinsic contribution of Rip2 in macrophages, we investigated signalling in macrophages infected

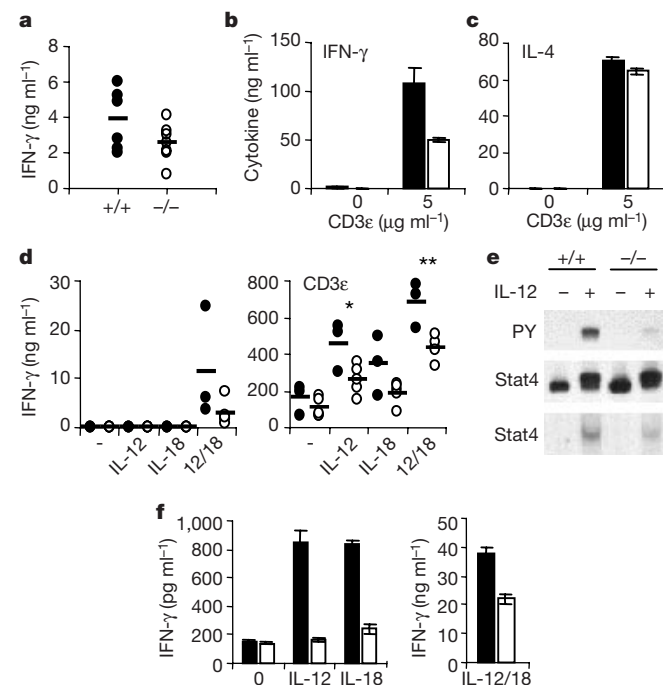


Figure 3 Impaired IL-12- and IL-18-induced activation of T_H1 and NK cells in *Rip2*^{-/-} mice. **a**, Lymph node cells from OVA (100 μ g) in CFA-immunized *Rip2*^{+/+} (filled circles) and *Rip2*^{-/-} (open circles) mice were cultured with OVA for 24 h before measuring IFN- γ production by ELISA. The horizontal bar indicates mean IFN- γ levels. **b**, **c**, Splenic CD4⁺ T cells of *Rip2*^{+/+} (filled circles) and *Rip2*^{-/-} (open circles) mice were cultured with anti-CD3 under T_H1-polarizing conditions (**b**) or under T_H2-polarizing conditions (**c**). Differentiated cells were restimulated with anti-CD3, and levels of cytokines were measured. **d**, T_H1 clones of *Rip2*^{+/+} (filled circles) and *Rip2*^{-/-} (open circles) mice were cultured with anti-CD3 as indicated, and assayed for IFN- γ production. Asterisk, $P < 0.04$ compared with *Rip2*^{+/+} cells; double asterisk, $P < 0.01$ compared with *Rip2*^{+/+} cells. **e**, T_H1 clones of *Rip2*^{+/+} and *Rip2*^{-/-} mice were stimulated with IL-12 as indicated. Immunoprecipitated Stat4 proteins were analysed by western analysis for phosphotyrosine (top circles), and for total Stat4 levels (middle circles). Activation of Stat4 was detected by EMSA (bottom). **f**, Splenocytes of *Rip2*^{+/+} (filled circles) and *Rip2*^{-/-} (open circles) mice were stimulated as indicated, then evaluated for IFN- γ production.

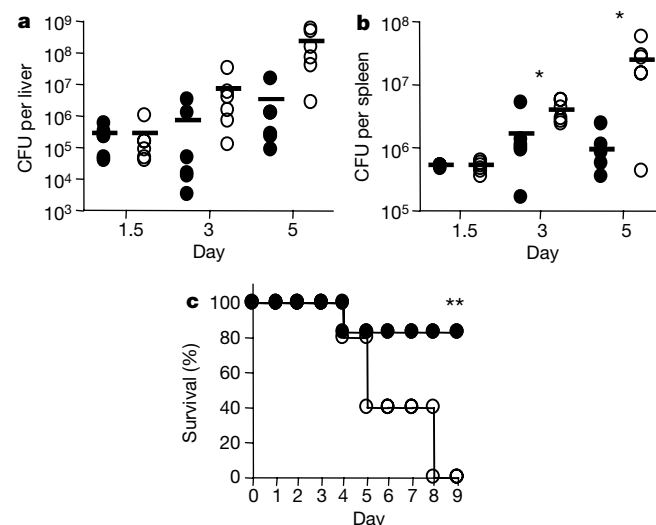


Figure 4 Increased susceptibility to *Listeria monocytogenes* in *Rip2*^{-/-} mice. **a**–**c**, *Rip2*^{+/+} (filled circles) and *Rip2*^{-/-} (open circles) mice were infected intravenously with 3×10^3 colony-forming units (CFU) of *L. monocytogenes*. 1.5, 3, or 5 days after infection, livers (**a**) and spleens (**b**) were collected and bacterial titres determined. Survival was determined after 9 days (**c**). Horizontal bars represent means. Asterisk, $P < 0.05$ compared with *Rip2*^{+/+} mice; double asterisk, $P < 0.01$ compared with *Rip2*^{+/+} mice.

with *L. monocytogenes*. Decreased activation of NF- κ B (Fig. 5a), a pathway critical to host inflammatory responses, was observed in macrophages of *Rip2*^{-/-} mice. Similar to infection with *L. monocytogenes*, LPS-stimulated macrophages of *Rip2*^{-/-} mice showed decreased activation of NF- κ B compared with *Rip2*^{+/+} cells (Fig. 5b), as well as decreased p38 and SAPK activation (data not shown). In addition, LPS stimulation upregulated *Rip2* expression as measured by RT-PCR (data not shown). To assess *in vivo* responsiveness, we challenged *Rip2*^{+/+} and *Rip2*^{-/-} mice with LPS, and found that *Rip2*^{-/-} mice were completely resistant to an LD₅₀ dose (Fig. 5c). This effect underscores the importance of *Rip2* in the recognition and/or subsequent activation of inflammatory responses to LPS, and highlights a common signalling defect in *Rip2*^{-/-} mice in TLR and related receptors such as IL-18.

Recent studies have implicated the plant disease-resistance-like proteins, Nod1 and Nod2, as activators of NF- κ B in response to presentation of pathogens in the cytosol¹⁵. In fact, mutations in Nod2 have been linked to susceptibility to Crohn's disease, implicating innate immune dysfunction in Crohn's disease^{16,17}. Considering how *Rip2*-deficient mice could respond *in vivo* with altered sensitivity to both LPS and *L. monocytogenes*, we investigated a role for *Rip2* in the signalling pathway of TLRs and Nods. Activation of NF- κ B by both TLR4 and Nod1 was inhibited in a dose-dependent manner by coexpression of dominant negative *Rip2* (*Rip2*-DN), as measured by a NF- κ B-dependent luciferase reporter in 293T cells (Fig. 5d). Whereas activation of NF- κ B by both TLR4 and Nod1 was inhibited using a dominant active inhibitor of NF- κ B (I κ B-DA), dominant negative forms of proteins found in TLR complexes (IRAK-DN and MyD88-DN) specifically abolished TLR4- but not

Nod1-induced NF- κ B activity. These data are consistent with previous studies demonstrating association of *Rip2* with either Nod1 or TRAF6, and further supports a *Rip2*-dependent contribution in TLR/IL-18 receptor and Nod1 family signalling^{18–20}.

Insight into the regulation of cell-mediated immunity is critical to understanding the balance between normal immune responses and inappropriate inflammatory responses. Our studies provide genetic evidence for the requirement of *Rip2* in both innate and adaptive immune responses. In the future, it will be important to establish the precise mechanisms of *Rip2* function in IL-12/Stat4 signalling, as well as in *in vivo* specificity in intracellular versus extracellular microbial infections. Here, we implicate *Rip2* in the innate response to pathogens by mediating Nod- and TLR-induced cell signalling. In the adaptive immune response, *Rip2* mediates cytokine-induced IFN- γ production in T_H1 and NK cells. The critical roles of *Rip2* in regulating cell-mediated responses provide support for *Rip2* as a candidate target for immune intervention. □

Methods

Generation of *Rip2*^{-/-} mice

An internal 1.1-kilobase (kb) *Hpa*I–*Sma*I genomic fragment containing exon 1 was replaced with a PGK-*neo*^r (neomycin-resistance gene) cassette. The targeting construct was electroporated into 129/J1 cells and selected with G418. Recombinants were identified using a genomic probe external to the targeting construct, hybridizing to an 8.8-kb *Bgl*II fragment of the endogenous *Rip2* locus and to an 11-kb fragment of the targeted allele. We used *Rip2*-specific C-terminal antibodies for western analysis (Alexis Biochemicals). Two clones chosen for injection into 3.5-day-old C57BL/6 blastocysts gave germ-line transmission of the targeted allele. Chimaeric mice were crossed with C57BL/6 females to generate *Rip2*^{+/-} mice. Mice were re-derived into a specific pathogen-free facility. Age- and gender-matched mice of C57BL/6x129 F₂ background were used, except for *L. monocytogenes* infections, in which mice of C57BL/6x129 F₃ generation were used.

PCR with reverse transcription

First strand complementary DNA was produced by reverse transcription (RT) of 5 μ g total RNA using the superscript first strand synthesis kit (Gibco BRL). Levels of hypoxanthine guanine phosphoribosyltransferase (HPRT) were amplified by PCR and used to equalize levels of cDNA²¹. T cells were differentiated *in vitro* by stimulation with anti-CD3 (5 μ g ml⁻¹) and anti-CD28 (1 μ g ml⁻¹) antibodies without antigen-presenting cells. The primers used for *Rip2* PCR were: 5'-CCATCCCGTACCACAAGCTC-3' and 5'-GCAGG ATGCGGAATCTCAAT-3'.

In vitro CD4⁺ T-cell differentiation

Splenic CD4⁺ T cells were purified by negative selection using a magnetic cell sorter (Stem Cell) to 90–95% purity as determined by staining with fluorescein isothiocyanate-conjugated anti-CD4 and flow cytometry. CD4⁺ depleted splenocytes were used as APCs. Purified T cells (5 $\times$ 10⁵ ml⁻¹) were cultured with wild-type irradiated APCs (5 $\times$ 10⁵ ml⁻¹) in the presence of anti-CD3 (5 μ g ml⁻¹) and IL-2 (20 U ml⁻¹). T_H1 polarization was achieved with the addition of IL-12 (5 ng ml⁻¹) and anti-IL-4 (4 μ g ml⁻¹), whereas for T_H2 polarization, IL-4 (1,000 U ml⁻¹) and anti-IFN- γ (4 μ g ml⁻¹) was added. After 4 days of culture, cells were collected, washed extensively, and restimulated at 5 $\times$ 10⁵ cells ml⁻¹ with plate-bound anti-CD3 (5 μ g ml⁻¹). Supernatants were collected after 24 h and assayed for levels of IFN- γ and IL-4 by ELISA. Antibodies for differentiation and ELISA assays were purchased from Pharmingen.

IFN- γ production in T_H1 clones and NK cells

T_H1 clones were maintained using irradiated wild-type APCs, and used for experiments 2 weeks after restimulation to decrease the contribution of cells from the APC population. T_H1 clones were cultured with media, IL-12 (1 ng ml⁻¹), IL-18 (50 ng ml⁻¹), or both, with and without plate-bound anti-CD3 (5 μ g ml⁻¹) for 48 h. Splenocytes were stimulated with media, IL-12 (5 ng ml⁻¹), IL-18 (20 ng ml⁻¹), or both for 24 h. FACS analysis of splenocytes using a NK-specific antibody in conjunction with intracellular IFN- γ staining demonstrated that IFN- γ secreted from splenocytes was derived from NK cells, and that similar numbers of NK cells were present in spleens of *Rip2*^{+/+} and *Rip2*^{-/-} mice. Supernatants were assessed for IFN- γ production by ELISA.

Analysis of Stat4 activation

T_H1 clones from *Rip2*^{+/+} and *Rip2*^{-/-} mice were stimulated with IL-12 (10 ng ml⁻¹) for 15 min. We prepared total cell extracts as described²². Stat4 immunoprecipitated with polyclonal anti-Stat4 antibodies (Santa Cruz) was assessed for phosphotyrosine with RC-20-horseradish peroxidase (Transduction Laboratories). After stripping, total Stat4 levels were detected using rabbit anti-Stat4 antibody. For DNA binding, lysates (20 μ g) were incubated with ³²P-labelled consensus binding site oligonucleotides for Stat4 (Santa Cruz), and analysed by electrophoresis mobility shift assay (EMSA).

NF- κ B activation in macrophages

Total bone marrow cells were isolated from mice and plated overnight (1 $\times$ 10⁶ cells ml⁻¹) in the presence of macrophage colony-stimulating factor (M-CSF). After 24 h,

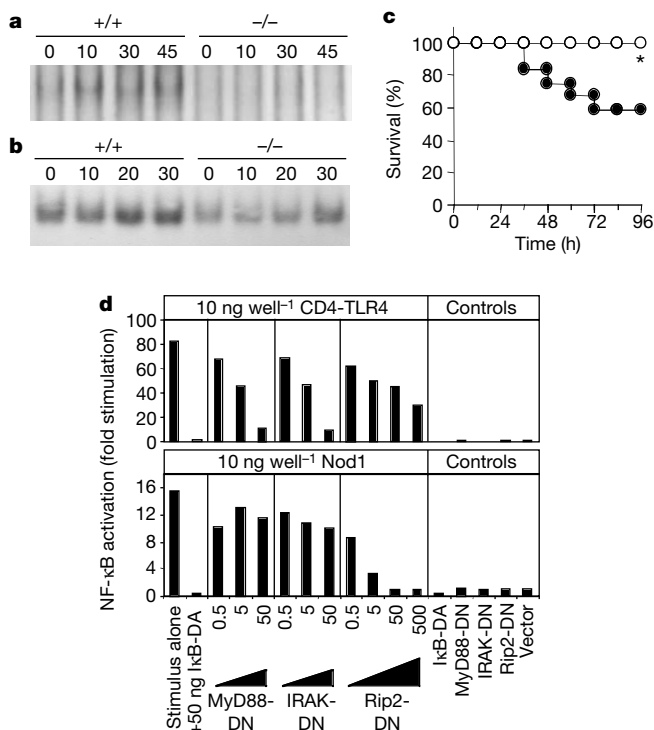


Figure 5 Impaired NF- κ B activation in macrophages of *Rip2*^{-/-} mice. **a**, **b**, Macrophages of *Rip2*^{+/+} and *Rip2*^{-/-} mice were stimulated with *L. monocytogenes* (**a**). IFN- γ -activated macrophages of *Rip2*^{+/+} and *Rip2*^{-/-} mice were stimulated with LPS (**b**). Activation of NF- κ B was determined by EMSA. **c**, Eleven *Rip2*^{+/+} (filled circles) and twelve *Rip2*^{-/-} (open circles) mice were intraperitoneally injected with 28 mg kg⁻¹ body weight of LPS from *Escherichia coli* serotype O55:B5, and monitored for survival. Asterisk, $P < 0.001$ compared with *Rip2*^{+/+} mice. **d**, Cell lysates from 293T cells transfected with the indicated constructs were analysed for luciferase activity of a NF- κ B reporter. Normalized fold activation is indicated.

non-adherent cells were plated ($1 \times 10^5 \text{ ml}^{-1}$) and cultured for 6 days. Unactivated macrophages were stimulated with *L. monocytogenes* using a multiplicity of infection of 20, whereas macrophages activated with IFN- γ (400 U ml^{-1}) for 16 h were stimulated with LPS. Nuclear extracts ($3 \mu\text{g}$) were incubated with a ^{32}P -labelled oligonucleotide probe derived from the major histocompatibility complex (MHC) class II NF- κB consensus-binding site, and analysed by EMSA.

Transient transfection and luciferase analysis

Human embryonic kidney (293T) cells were plated in 6-well plates at 2.5×10^5 cells per well. Cells were stimulated by transfection with plasmid pCMV (cytomegalovirus) human NOD1 (10 ng) or pCMV murine CD4-TLR4 (10 ng). Stimulated cells were inhibited by co-transfection with pCMV human RIP2-DN (1–434), pCDNA3 murine MyD88-DN, pCiNeo murine IRAK-DN, or pCMV I κ B-DA, as indicated (in ng). All transfections were performed with a luciferase reporter plasmid containing two κB binding sites (20 ng) and pCMV-LacZ (40 ng). Luciferase activity was measured and normalized to beta-galactosidase activity.

Received 22 November 2001; accepted 10 January 2002.

- Medzhitov, R. & Janeway, C. A. Jr. Innate immune recognition and control of adaptive immune responses. *Semin. Immunol.* **10**, 351–353 (1998).
- McCarthy, J. V., Ni, J. & Dixit, V. M. RIP2 is a novel NF- κB -activating and cell death-inducing kinase. *J. Biol. Chem.* **273**, 16968–16975 (1998).
- Inohara, N., del Peso, L., Koseki, T., Chen, S. & Núñez, G. RICK, a novel protein kinase containing a caspase recruitment domain, interacts with CLARP and regulates CD95-mediated apoptosis. *J. Biol. Chem.* **273**, 12296–12300 (1998); erratum *J. Biol. Chem.* **273**, 18675 (1998).
- Thome, M. *et al.* Identification of CARDIAK, a RIP-like kinase that associates with caspase-1. *Curr. Biol.* **8**, 885–888 (1998).
- Yip, H. C. *et al.* Adjuvant-guided type-1 and type-2 immunity: infectious/noninfectious dichotomy defines the class of response. *J. Immunol.* **162**, 3942–3949 (1999).
- Finkelman, F. D. *et al.* Lymphokine control of *in vivo* immunoglobulin isotype selection. *Annu. Rev. Immunol.* **8**, 303–333 (1990).
- Mosmann, T. R. & Coffman, R. L. TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties. *Annu. Rev. Immunol.* **7**, 145–173 (1989).
- Paul, W. E. & Seder, R. A. Lymphocyte responses and cytokines. *Cell* **76**, 241–251 (1994).
- Yang, J., Murphy, T. L., Ouyang, W. & Murphy, K. M. Induction of interferon- γ production in TH1 CD4 $^{+}$ T cells: evidence for two distinct pathways for promoter activation. *Eur. J. Immunol.* **29**, 548–555 (1999).
- Thierfelder, W. E. *et al.* Requirement for Stat4 in interleukin-12-mediated responses of natural killer and T cells. *Nature* **382**, 171–174 (1996).
- Kaplan, M. H., Sun, Y. L., Hoey, T. & Grusby, M. J. Impaired IL-12 responses and enhanced development of Th2 cells in Stat4-deficient mice. *Nature* **382**, 174–177 (1996).
- Trinchieri, G. Interleukin-12: a proinflammatory cytokine with immunoregulatory functions that bridge innate resistance and antigen-specific adaptive immunity. *Annu. Rev. Immunol.* **13**, 251–276 (1995).
- Dai, W. J. *et al.* Impaired macrophage listericidal and cytokine activities are responsible for the rapid death of *Listeria monocytogenes*-infected IFN- γ receptor-deficient mice. *J. Immunol.* **158**, 5297–5304 (1997).
- DiTirro, J. *et al.* Disruption of the cellular inflammatory response to *Listeria monocytogenes* infection in mice with disruptions in targeted genes. *Infect. Immun.* **66**, 2284–2289 (1998).
- Inohara, N., Ogura, Y., Chen, F. E., Muto, A. & Núñez, G. Human Nod1 confers responsiveness to bacterial lipopolysaccharides. *J. Biol. Chem.* **276**, 2551–2554 (2001).
- Hugot, J. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603 (2001).
- Ogura, Y. *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**, 603–606 (2001).
- Inohara, N. *et al.* An induced proximity model for NF- κB activation in the Nod1/RICK and RIP signaling pathways. *J. Biol. Chem.* **275**, 27823–27831 (2000).
- Ogura, Y. *et al.* Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF- κB . *J. Biol. Chem.* **276**, 4812–4818 (2001).
- Girardin, S. E. *et al.* CARD4/Nod1 mediates NF- κB and JNK activation by invasive *Shigella flexneri*. *EMBO Rep.* **2**, 736–742 (2001).
- Reiner, S. L., Zheng, S., Corry, D. B. & Locksley, R. M. Constructing polycompetitor cDNAs for quantitative PCR. *J. Immunol. Methods* **165**, 37–46 (1993); erratum *J. Immunol. Methods* **173**, 133; **175**, 275 (1994).
- Khan, K. D. *et al.* Induction of the Ly-6A/E gene by interferon α/β and gamma requires a DNA element to which a tyrosine-phosphorylated 91-kDa protein binds. *Proc. Natl Acad. Sci. USA* **90**, 6806–6810 (1993).

Acknowledgements

We thank V. Dixit for human RIP2 cDNA; T. Parks for Nod1 cDNA; T. Roni and S. Smale for CD4-TLR4 cDNA; K. Shau and X.-F. Qin for reagents and advice; and R. L. Modlin for critical reading of the manuscript. A.I.C. is supported by the Medical Scientist Training Program; P.W.D. is a Lymphoma and Leukemia Society Research Fellow; and G.C. is a Lymphoma and Leukemia Society Research Scholar.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.C. (e-mail: genhongc@microbio.ucla.edu). The sequence for murine Rip2 cDNA has been deposited in GenBank under accession number AF461040.

RICK/Rip2/CARDIAK mediates signalling for receptors of the innate and adaptive immune systems

Koichi Kobayashi*, Naohiro Inohara†, Lorraine D. Hernandez‡, Jorge E. Galán‡, Gabriel Núñez†, Charles A. Janeway*§, Ruslan Medzhitov*§ & Richard A. Flavell*§

* Section of Immunobiology, and the § Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, Connecticut 06520, USA

† Department of Pathology and Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA

‡ Section of Microbial Pathogenesis, Boyer Center for Molecular Medicine, Yale School of Medicine, New Haven, Connecticut 06536, USA

The immune system consists of two evolutionarily different but closely related responses, innate immunity and adaptive immunity. Each of these responses has characteristic receptors—Toll-like receptors (TLRs) for innate immunity and antigen-specific receptors for adaptive immunity. Here we show that the caspase recruitment domain (CARD)-containing serine/threonine kinase Rip2 (also known as RICK, CARDIAK, CCK and Ripk2)^{1–4} transduces signals from receptors of both immune responses. Rip2 was recruited to TLR2 signalling complexes after ligand stimulation. Moreover, cytokine production in Rip2-deficient cells was reduced on stimulation of TLRs with lipopolysaccharide, peptidoglycan and double-stranded RNA, but not with bacterial DNA, indicating that Rip2 is downstream of TLR2/3/4 but not TLR9. Rip2-deficient cells were also hyporesponsive to signalling through interleukin (IL)-1 and IL-18 receptors, and deficient for signalling through Nod proteins—molecules also implicated in the innate immune response. Furthermore, Rip2-deficient T cells showed severely reduced NF- κB activation, IL-2 production and proliferation on T-cell-receptor (TCR) engagement, and impaired differentiation to T-helper subtype 1 (TH1) cells, indicating that Rip2 is required for optimal TCR signalling and T-cell differentiation. Rip2 is therefore a signal transducer and integrator of signals for both the innate and adaptive immune systems.

Rip2 is a serine/threonine kinase^{1–4} that carries a CARD at its carboxy terminus, and shares sequence similarity with a serine/threonine kinase, Rip, which is essential for NF- κB activation through the tumour-necrosis factor (TNF) receptor⁵. *In vitro* studies have shown that Rip2 can associate with a variety of other CARD-containing molecules through CARD–CARD interactions^{1–3}. Moreover, overexpression of Rip2 causes activation of NF- κB and Jun amino-terminal kinase (JNK)^{1–4}. NF- κB activation by Rip2 is inhibited by dominant negative TRAF6 (ref. 4), a signalling molecule that is downstream of TLRs. Furthermore, we found that expression of Rip2 was induced in macrophages on stimulation with lipopolysaccharide (LPS) (Fig. 1a, b). These observations led us to consider the possibility that Rip2 is involved in signalling in the innate immune system. To assess the physiological role of Rip2 in signalling in the innate immune system, we generated Rip2-deficient mice by homologous recombination of embryonic stem cells. A gene-targeting construct was generated to replace the two exons coding for murine *Rip2* with a neomycin-resistance gene (*neo*^r) (Fig. 1c). Homologous recombination in embryonic stem cells was confirmed by Southern blot analysis (Fig. 1d), and the absence of Rip2 expression in homozygous animals was confirmed by western blot (Fig. 1e). Rip2-deficient mice were born in the expected mendelian ratio, and showed no gross developmental abnormalities and no abnormal composition of lymphocytes as determined by flow cytometry (data not shown).

TLRs can recognize specific pathogen-associated molecular patterns (PAMPs) such as LPS⁶, lipoteichoic acid (LTA)⁷, peptidoglycan (PGN)⁷, CpG-containing DNA⁸, or double-stranded RNA⁹. To test whether Rip2 is involved in TLR signalling, macrophages of *Rip2*^{-/-} mice were stimulated with various PAMPs, and cytokine/chemokine production was assessed by enzyme-linked immunosorbent assay (ELISA). Production of the inflammatory cytokines IL-6 and TNF- α , and the chemokine IP10, was severely reduced in macrophages of *Rip2*^{-/-} mice on stimulation with LPS and LTA (both of which are ligands for TLR4), or PGN (which is a ligand for TLR2) (Fig. 2a). There was no defect in cytokine/chemokine production after stimulation with CpG DNA (a ligand for TLR9), indicating that Rip2 is involved in TLR4 and TLR2 but not TLR9 signalling. Production of IL-6 is also reduced in *Rip2*^{-/-} embryonic fibroblasts on stimulation with double-stranded RNA, poly(IC) (a ligand for TLR3) and LPS in a dose- and time-dependent manner (Fig. 2c). These results indicate that Rip2 is required for normal levels of signalling through some, but not all, TLRs. To assess the response to a live pathogen, macrophages of *Rip2*^{-/-} mice were infected with *Listeria monocytogenes*, and the levels of IL-6 and TNF- α were measured by ELISA. Macrophages of *Rip2*^{-/-} mice were compromised in their ability to produce these cytokines, further supporting the involvement of this kinase in the innate immune response (Fig. 2b and data not shown). To examine the mechanism of activation of Rip2 in macrophages infected with *L. monocytogenes*, in particular to determine whether stimulation of the innate immune response by *Listeria* occurs at the cell surface or intracellularly, cytochalasin D was added to cell cultures at various concentrations, as it blocks *Listeria* internalization¹⁰. IL-6 production in response to *Listeria* infection was not altered even at high concentrations of cytochalasin D, and in all cases was reduced in *Rip2*^{-/-} cells (Fig. 2b). Thus, attachment of bacteria to the cell surface is sufficient to activate macrophages, and the reduced cytokine production in *Rip2*^{-/-} cells is due to defective signalling

from cell-surface receptors. *In vivo* response to LPS by *Rip2*-deficient mice was assessed by endotoxin shock experiments using intraperitoneal injection of LPS. *Rip2*-deficient mice were more resistant to LPS than wild-type mice (Fig. 2d), suggesting the importance of this molecule in LPS response *in vivo*.

Mature IL-1 β can be produced through cleavage of pro-IL-1 β by active caspase-1. Previous *in vitro* studies have suggested that Rip2 binds and activates pro-caspase-1, thereby generating IL-1 β ^{3,11}. Notably, macrophages of *Rip2*^{-/-} mice exhibited no reduction in IL-1 β production when cells were stimulated with LPS, CpG DNA and peptidoglycan for 3 h (Supplementary Information), suggesting that Rip2 is not involved in caspase-1 activation downstream of TLRs, or alternatively that there is another redundant mechanism.

TLR signalling requires the formation of multiprotein signalling complexes, which include the serine/threonine kinase IRAK and the adapter molecules MyD88 and TRAF6, resulting in the activation of

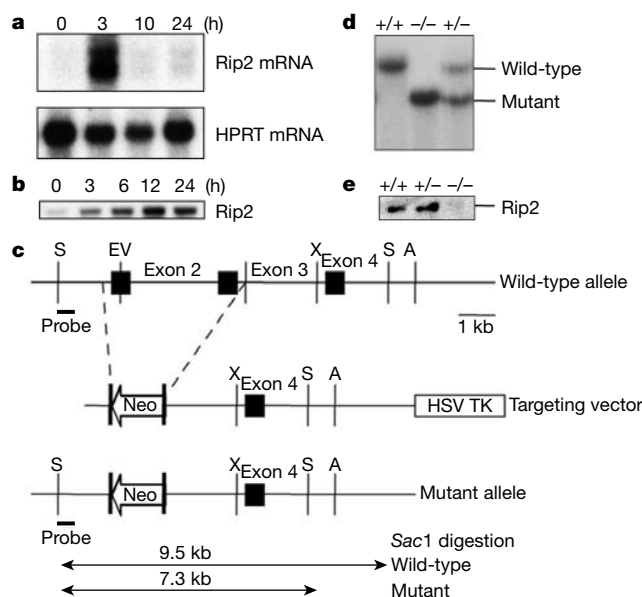


Figure 1 Expression of Rip2 in macrophages and targeted disruption of the mouse *Rip2* gene. **a**, **b**, Northern and Western blot analysis for Rip2 expression in macrophages stimulated with 10 ng ml⁻¹ LPS for the indicated times. HPRT, hypoxanthine guanine phosphoribosyltransferase. **c**, Diagram of the *Rip2* locus, the targeting vector and the targeted allele. Filled boxes denote exons. Restriction enzyme sites are indicated (S, *SacI*; EV, *EcoRV*; X, *XbaI*; A, *ApaI*). TK, thymidine kinase. **d**, Southern blot analysis of *SacI*-digested genomic DNA identifies mice of all three expected genotypes. **e**, Western blot analysis of thymocytes showing the absence of Rip2 protein in homozygous mice.

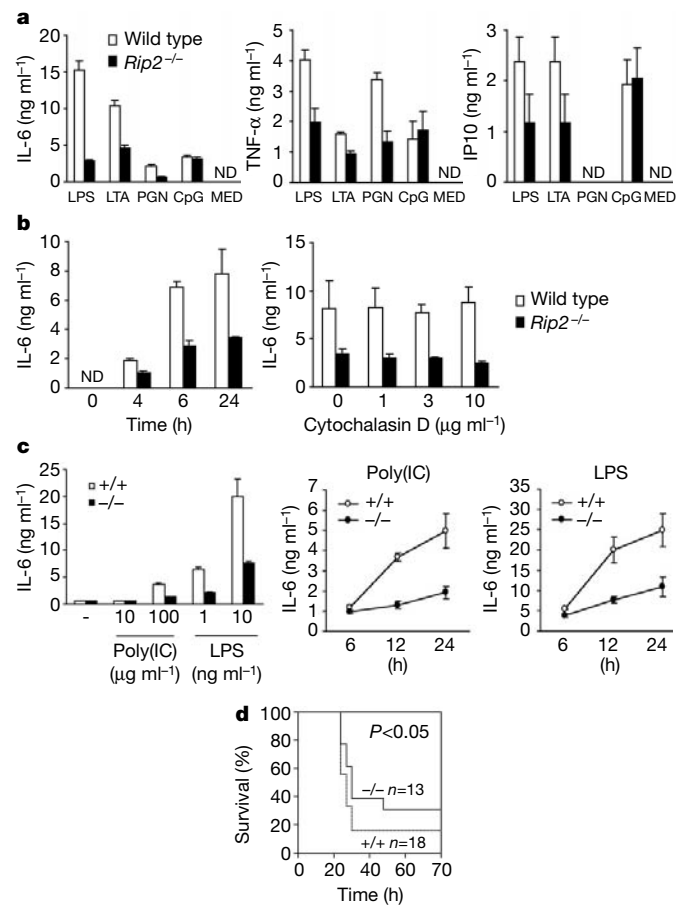


Figure 2 Impaired TLR responses in *Rip2*-deficient cells. **a**, Production of IL-6, TNF- α and IP10 by *Rip2*^{-/-} macrophages derived from bone marrow, stimulated with LPS (10 ng ml⁻¹), LTA (10 μ g ml⁻¹), PGN (10 μ g ml⁻¹), CpG oligonucleotide DNA (CpG, 10 μ M), and medium alone (MED) for 6 h. ND, not detected. **b**, Production of IL-6 by *Rip2*^{-/-} macrophages infected with *L. monocytogenes* and cultured for the indicated period (left), or cultured for 6 h in the presence of the indicated concentrations of cytochalasin D (right). **c**, Production of IL-6 by *Rip2*^{-/-} fibroblasts stimulated with poly(IC) or LPS with indicated dose for 24 h (left), or stimulated with 100 μ g ml⁻¹ poly(IC) (middle) or 10 ng ml⁻¹ LPS (right) for indicated periods. **d**, Survival curve of wild-type and *Rip2*^{-/-} mice after endotoxin shock. LPS (16.7 mg kg⁻¹ body weight) was injected intraperitoneally into wild-type and *Rip2*^{-/-} mice. Viability was determined every 3 h for the first 36 h, followed by every 12 h. There was no incremental death after 80 h. *P*-values were determined by the Mantel–Cox test. Experiments were performed with triplicates and were repeated 6 (**a**, IL-6), 3 (**a**, TNF- α /IP10) and 2 (**b**, **c**) times, with equivalent results.

NF- κ B and the mitogen-activated protein (MAP) kinases JNK, p38 and ERK1/2 (refs 12, 13). We therefore studied activation of these molecules in macrophages of *Rip2*^{-/-} mice by examining their phosphorylation state. LPS-stimulated *Rip2*^{-/-} macrophages showed reduced phosphorylation levels of p38, I κ B α , ERK and JNK, and reduced degradation of I κ B α (Fig. 3a), indicating altered signalling downstream from TLR4. This reduced signalling was not due to changes in the expression levels of MyD88, IRAK or TRAF6, as western blot analysis for these proteins showed no difference between wild-type and *Rip2*^{-/-} macrophages (data not shown). Furthermore, we studied whether Rip2 is recruited to TLR signalling complexes. 293T cells stably transfected with Flag-tagged TLR2 and CD14 (Flag-TLR2/CD14 293T cells) were stimulated with PGN, and association of TLR2 and endogenous Rip2 was examined by immunoprecipitation using an anti-Flag monoclonal antibody. Association of Rip2 with TLR2 was induced by PGN and peaked at 10 min after stimulation; the amount of associated Rip2 decreased after 15 min, indicating that the recruitment of Rip2 to TLR signalling complex is transient after TLR activation (Fig. 3b).

In addition to TLRs, increasing evidence implicates another family of proteins in innate immune responses. The cytoplasmic proteins collectively termed Nod, are characterized by the presence of three motifs: a CARD, a nucleotide-binding domain (NBD) and a leucine-rich repeat (LRR) region. These proteins have homology to the NBD-LRR-like disease resistant proteins in plants^{14–16}. An increasing number of members of this family have been identified (Nod1/CARD4, Nod2, DEFCAP/NAC, CARD12/Ipaf/CLAN), and by analogy to the plant molecules these data imply that, like the TLR family, Nod proteins are a diverse family of molecules designed to detect pathogens in intracellular compartments—the LRR region of members of both families probably confers pathogen specificity^{17,18}. In fact, Nod1 is activated on infection of *Shigella flexneri* in epithelial cells¹⁹, and one NBD-LRR protein, NAIP, determines susceptibility to *Legionella pneumophila* infection²⁰. Also, it has been demonstrated that Nod2 is mutated in patients susceptible to Crohn's disease and Blau syndrome^{18,21,22}. *In vitro* studies have revealed that Nod1 and Nod2 bind to Rip2 by means of a CARD–CARD interaction^{15,16}, suggesting that Rip2 may be involved in signalling downstream of the Nod family proteins. To test this, we generated Rip2-deficient embryonic fibroblasts and co-transfected them with both Nod expression vectors and a NF- κ B reporter construct. NF- κ B activation by Nod1/Nod2 expression was completely abolished in fibroblasts of *RIP2*^{-/-} mice, and complementation of *Rip2*^{-/-} fibroblasts with a Rip2 expression vector restored these defects (Fig. 4). These results indicate that Rip2 is essential for

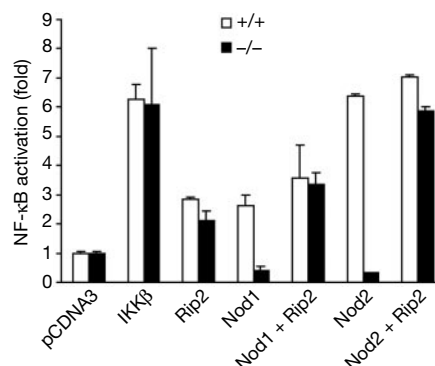


Figure 4 Rip2 is essential for activation of NF- κ B by Nod1 and Nod2. Embryonic fibroblasts from wild-type and Rip2-deficient mice were co-transfected with pcDNA3, pcDNA3-IKK β , pcDNA3-Nod1 Δ LRR, pcDNA3-Nod2 or pcDNA3-Rip2, pEF-BOS- β -gal and pBVI-Luc reporter plasmids. The experiments were repeated at least three times with equivalent results.

NF- κ B activation by both Nod1 and Nod2. Together, these results indicate that Rip2 is essential for signalling through both protein families of the innate immune system, that is TLRs and Nod.

IL-1 and IL-18 receptors both have a TIR (Toll/IL1-receptor) domain within the cytoplasmic tail, and therefore belong to the same multigene family as the TLRs. We therefore tested whether responses to IL-1 or IL-18 were altered in cells of *Rip2*^{-/-} mice. Under certain conditions, IL-1 β is a potent co-stimulant to T cells for their growth. We stimulated thymocytes with IL-1 β , together with IL-2 or a low dose of Concanavalin A (Con A). Thymocytes of *Rip2*^{-/-} mice showed reduced proliferation upon IL-1 β stimulation in both IL-2 and Con A co-stimulation (Fig. 5a). Embryonic fibroblasts produce IL-6 upon stimulation with the cytokines IL-1 β or TNF- α . We found that embryonic fibroblasts of *Rip2*^{-/-} mice were significantly impaired in their ability to produce IL-6 when stimulated with IL-1 β , but not when stimulated with TNF- α , suggesting that Rip2 is involved in IL-1- but not TNF- α -receptor signalling (Fig. 5b).

We next investigated the IL-18 response in cells of *Rip2*^{-/-} mice. IFN- γ production by natural killer (NK) cells upon IL-18 stimulation was assessed using splenocytes of *Rip2*^{-/-} mice. IFN- γ production was severely reduced in *Rip2*^{-/-} cells by stimulation with IL-18 (Fig. 5c). Surprisingly, IFN- γ production upon stimulation with IL-12 was also reduced. Co-stimulation of *Rip2*^{-/-} cells with IL-18 and IL-12 also resulted in reduced production of IFN- γ

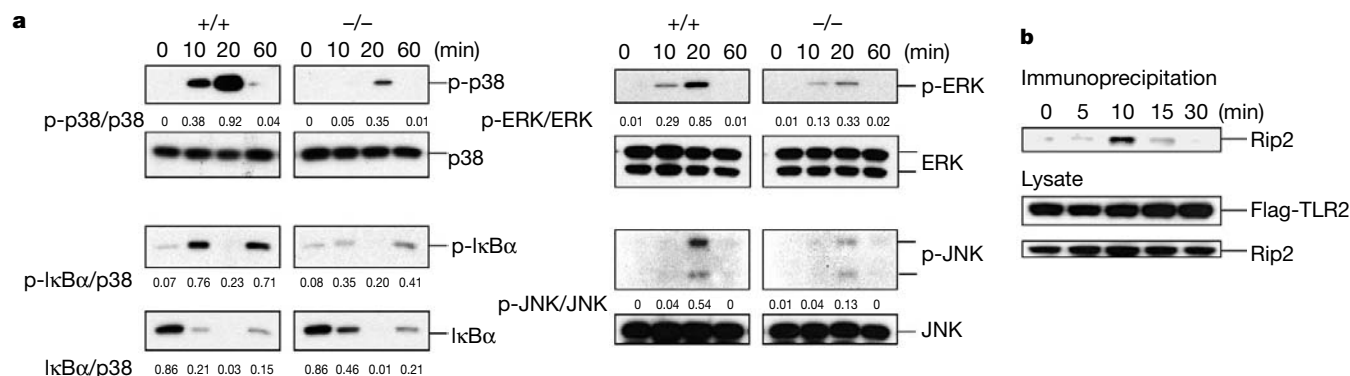


Figure 3 Rip2 is required for optimal TLR signalling. **a**, Rip2 is upstream of multiple signalling pathways, including NF- κ B, JNK, p38 and ERK. Macrophages derived from bone marrow of wild-type and Rip2-deficient mice were stimulated with LPS (10 ng ml⁻¹) for indicated periods. Cell lysates were prepared and blotted with the indicated antibodies. Ratios of phosphorylated and unphosphorylated proteins quantified by an image-analyser

are shown. These data are the mean of three independent experiments. **b**, Transient recruitment of Rip2 to TLR signalling complexes. Flag-TLR2/CD14 293T cells were stimulated with 10 μ g ml⁻¹ PGN for the indicated times. Cell lysates were immunoprecipitated with anti-Flag monoclonal antibody and blotted with anti-Rip2 antibody.

(Fig. 5c). Furthermore, we examined IL-18 and IL-12 response using differentiated effector CD4⁺ T_H1 cells. Similar to NK cells, effector T_H1 cells make copious amounts of IFN- γ when stimulated with IL-18 and IL-12 in the absence of T-cell receptor (TCR) stimulation²³. IFN- γ production by *Rip2*^{-/-} T_H1 cells stimulated with IL-18, IL-12 or a combination of IL-18 and IL-12, was severely perturbed (Fig. 5d). These results suggest that Rip2 is involved in

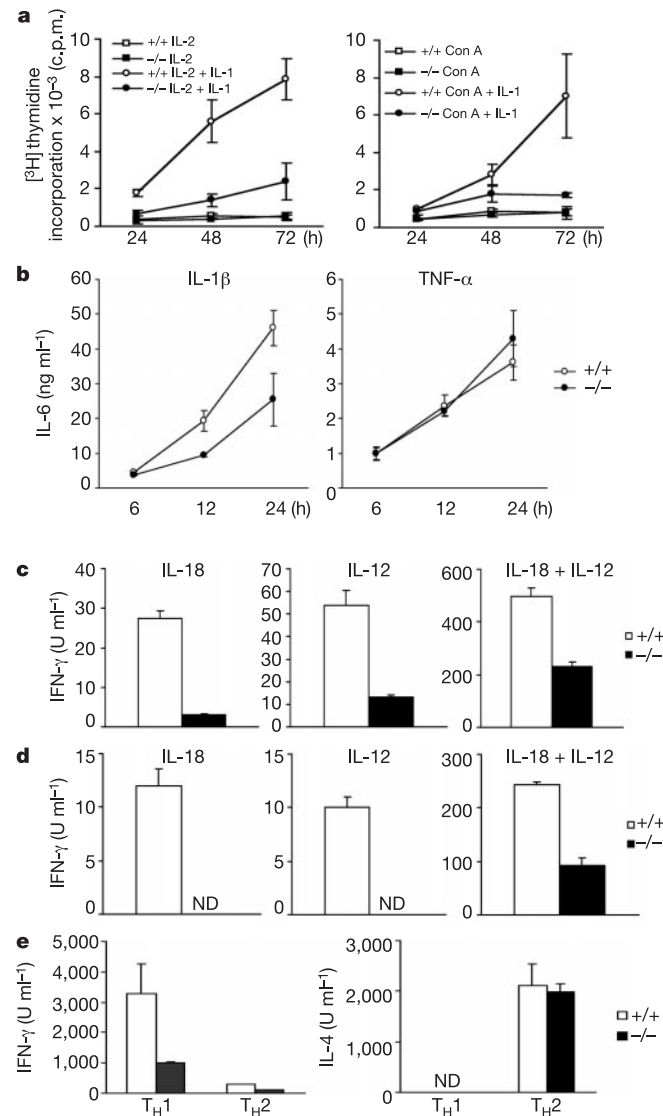


Figure 5 Rip2 is required for optimal IL-1/IL-18 receptor signalling in NK cells and T_H1 effector CD4 T cells. **a**, Proliferation of *Rip2*^{-/-} thymocytes stimulated with IL-2 (2 ng ml⁻¹) or Con A (0.625 μ g ml⁻¹) alone or together with IL-1 β (10 ng ml⁻¹), and cultured for indicated periods. Cells were pulsed with [³H]thymidine 8 h before collection. **b**, IL-6 production by *Rip2*^{-/-} embryonic fibroblasts stimulated with IL-1 β (left, 10 ng ml⁻¹) or TNF- α (right, 10 ng ml⁻¹) for indicated periods. **c**, Production of IFN- γ by NK cells on stimulation with IL-18 and IL-12. Splenocytes from wild-type or *Rip2*^{-/-} mice were stimulated with IL-18 (10 ng ml⁻¹) or IL-12 (10 ng ml⁻¹) alone for 48 h, or with both IL-18 (10 ng ml⁻¹) and IL-12 (1 ng ml⁻¹) for 24 h. **d**, Production of IFN- γ by T_H1 cells stimulated with either IL-18 (10 ng ml⁻¹) or IL-12 (10 ng ml⁻¹) alone, or with both IL-18 and IL-12 for 24 h. **e**, T_H1/T_H2 differentiation of CD4⁺ T cells *in vitro*. CD4⁺ T cells from wild-type or *Rip2*^{-/-} mice were stimulated with 10 μ g ml⁻¹ of plate-bound anti-CD3, and cultured for 4 days under T_H1 conditions (3.5 ng ml⁻¹ IL-12 and 2 μ g ml⁻¹ anti-IL-4 antibody) or T_H2 conditions (1,000 U ml⁻¹ IL-4 and 1 μ g ml⁻¹ anti-IFN- γ antibody). After washing, cells were counted and re-stimulated with 10 μ g ml⁻¹ anti-CD3 and cultured for 24 h. All experiments were performed in triplicates and were repeated three times, with equivalent results.

signalling downstream of the TLR/IL-1 receptor family, and the altered IL-12 response suggests that Rip2 may be involved in IL-12 signalling either directly or indirectly. IL-12 is one of the principal cytokines regulating T-cell differentiation. We therefore analysed T-cell differentiation of *Rip2*^{-/-} T cells. CD4⁺ T cells were cultured for 4 days under either T_H1 or T_H2 conditions, and re-stimulated with plate-bound anti-CD3 antibodies for 24 h. IFN- γ production by T_H1 cells of *Rip2*^{-/-} mice was reduced, although IL-4 production of *Rip2*^{-/-} T_H2 cells was not affected. These results suggest that Rip2 has an important role in the differentiation of T_H1 cells but not T_H2 cells.

Altered TCR signalling has a profound influence on T_H1/T_H2 differentiation²⁴. We therefore examined the response of *Rip2*^{-/-} T cells on TCR stimulation with anti-CD3 antibodies. CD4⁺ T cells of *Rip2*^{-/-} mice showed severely reduced proliferation upon anti-CD3 stimulation in a dose- and time-dependent manner (Fig. 6a). IL-2 production was reduced in *Rip2*^{-/-} CD4⁺ T cells, and this defect could not be rescued by co-stimulation with anti-CD28 (Fig. 6b). As Rip2 is involved in NF- κ B activation¹⁻⁴, and that NF- κ B activation is required for T-cell proliferation upon TCR stimulation^{25,26}, we analysed NF- κ B activation in *Rip2*^{-/-} T cells upon anti-CD3 stimulation. Phosphorylation and degradation of I κ B α was assessed by western blotting, and was reduced in *Rip2*^{-/-} T cells (Fig. 6c). In addition, NF- κ B activation assessed by gel-shift assay was also substantially reduced (Fig. 6d). Taken together, these data indicated that Rip2 is required for optimal activation of NF- κ B and T-cell proliferation upon TCR stimulation. As inhibition of NF- κ B activation can cause T_H1 deficiency *in vivo*²⁷, the T_H1 deficiency in *Rip2*^{-/-} T cells may be attributable, at least in part, to altered TCR signalling and NF- κ B activation.

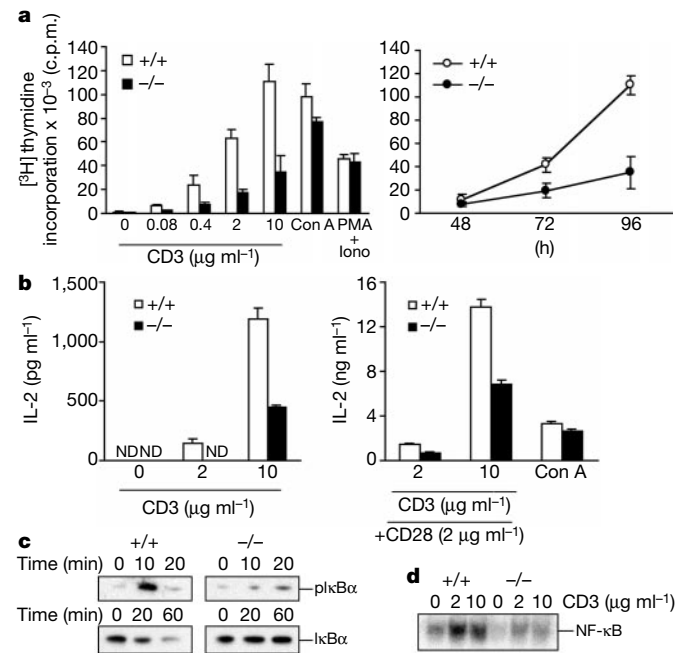


Figure 6 Rip2 is required for optimal NF- κ B activation, IL-2 production and T-cell proliferation on TCR stimulation. **a**, Proliferation of *Rip2*^{-/-} CD4⁺ T cells stimulated with Concanavalin A (Con A, 2.5 μ g ml⁻¹), PMA (40 ng ml⁻¹) plus ionomycin (0.5 μ M) or anti-CD3 at the indicated dose (96 h), or with 10 μ g ml⁻¹ anti-CD3 for indicated periods in the presence of irradiated T-cell-depleted splenocytes. **b**, IL-2 production by CD4⁺ T cells stimulated with plate-coated anti-CD3 either in the absence or presence of anti-CD28 (2 μ g ml⁻¹), or with Con A (2.5 μ g ml⁻¹). **c**, Phosphorylation and degradation of I κ B α in splenic T cells stimulated with anti-CD3 (10 μ g ml⁻¹), examined by western blotting. **d**, NF- κ B activation of splenic T cells stimulated with anti-CD3 with indicated dose for 8 h, analysed by gel mobility shift assay. Experiments in **a** and **b** were performed in triplicates and were repeated three times with equivalent results.

Our results indicate that Rip2 is required for signalling through both TLRs and Nod protein family members, which are central components of the innate immune system. We also show that Rip2 is required for appropriate TCR signalling. Therefore, Rip2 is a unique kinase capable of integrating signals emanating from both the innate as well as the adaptive immune systems. In fact, a notable feature of Rip2 deficiency is that it reveals defects in signalling from various immune receptors. There are three conceivable mechanistic explanations for this. The first is that Rip2 is a direct downstream signalling molecule of each receptor. Second, it is possible that Rip2 always receives signals from Nod proteins, and that perhaps different Nod proteins are the common element downstream of these different innate and adaptive immune receptor systems, and upstream of Rip2. The third possibility is that some receptors are dependent on Rip2 via Nod proteins for signalling, and others are not. Studies involving Nod-deficient mice should help to resolve this mechanistic issue. □

Methods

Generation of Rip2-deficient mice

A murine partial complementary DNA that codes for Rip2 was obtained by polymerase chain reaction, using mouse heart first-strand cDNA (Clontech) as the template and a specific primer based on GenBank accession number AA655189 and a primer based on the human *RIP2* sequence. A 129SV/J genomic library (Stratagene) was screened with the murine *Rip2* cDNA, and two phage carrying overlapping genomic clones encompassing *Rip2* were isolated. A targeting vector was designed to replace a 4.0-kilobase genomic fragment containing the second and third exons coding for the active site aspartate residue with the *loxP*-flanked *neo^r* gene expression cassette. The targeting vector was linearized with *NotI* and electroporated into W9.5 embryonic stem cells. Clones resistant to G418 and gancyclovir were selected, and homologous recombination (6 out of 135 clones) was confirmed by Southern blotting. Three clones homologous for the targeted mutation were injected into C57BL/6 blastocysts, which were subsequently transferred into pseudopregnant foster mother mice.

Plasmids and reagents

The expression vectors pcDNA3-IKK β , pcDNA3-Myc-Rip2, pcDNA3-Nod1-Flag, pcDNA3-Nod2-Flag, pEF1-BOS- β -gal and pBVI-Luc were described previously¹⁶. Lipopolysaccharide (LPS) from *Salmonella abortus equi* and lipoteichoic acid (LTA) from *Staphylococcus aureus* were purchased from Sigma. Peptidoglycan (PGN) from *S. aureus* was from Fluka. Poly(IC) was from Amersham Pharmacia Biotech. Phosphorothioate-modified CpG oligonucleotide DNA (TCCATGACGTTCTCTGACGTT) was synthesized in the Howard Hughes Medical Institute Biopolymer and W. M. Keck Biotechnology Resource Laboratory in Yale University. Human IL-1 β and mouse TNF- α were from R&D mouse IL-18 was from MBL; mouse IL-12 was from Genetics Institute; mouse IL-4 was from PharMingen; and anti-CD3, anti-IL-4 or anti-IFN- γ were purified from supernatants of 2C11, 11B11 or XMG hybridoma, respectively.

Culture of bone-marrow-derived macrophages

Macrophages derived from bone marrow were prepared as described²⁸. Cells were collected with cold DPBS, washed, re-suspended in DMEM, supplemented with 10% fetal calf serum, and used at a density of $2 \times 10^5 \text{ ml}^{-1}$ in the experiments. Cells were left untreated for at least 4 h at 37°C in 10% CO₂ before further handling.

Listeria infection of macrophages

The cells were cultured without antibiotics, and *L. monocytogenes* (American Type Culture Collection strain 43251) was added at a multiplicity of infection (MOI) of 50. After incubation for 30 min, extracellular bacteria were removed by washing three times with DPBS. To prevent reinfection, we cultured the cells in medium containing gentamicin sulphate ($50 \mu\text{g ml}^{-1}$, Gibco BRL).

Measurement of cytokine production

Macrophages derived from bone marrow were cultured with the indicated concentration of LPS, LTA, PGN or CpG DNA for 6 h. Embryonic fibroblasts were cultured with poly(IC), LPS, IL-1 β or TNF- α at the indicated concentration for the indicated periods. The concentration of IL-6, TNF- α and IP10 in culture supernatants was measured by ELISA.

Proliferation assays

We stimulated thymocytes with IL-2 (2 ng ml^{-1}) or Con A ($0.625 \mu\text{g ml}^{-1}$), in the presence or absence of IL-1 β (10 ng ml^{-1}). Purified CD4⁺ T cells were stimulated with anti-CD3 (2C11) in the presence of T-cell-depleted irradiated splenocytes. Cells were pulsed with [³H]thymidine for 8 h, and its incorporation was measured with a β -plate counter (Wallac).

Cytokine production by NK cells and T cells

T_H1/T_H2 differentiated cells were washed and re-stimulated with $10 \mu\text{g ml}^{-1}$ anti-CD3, and cultured for 24 h. IFN- γ and IL-4 in the supernatant was measured. Total splenocytes

and T_H1 cells were stimulated with IL-18 (10 ng ml^{-1}), IL-12 (10 ng ml^{-1}) or a combination of both. We measured the concentration of IFN- γ and IL-4 by ELISA. For IL-2 production, purified CD4⁺ T cells were stimulated with plate-bound anti-CD3 in the presence or absence of anti-CD28 ($2 \mu\text{g ml}^{-1}$) at the indicated concentration for 24 h, and the level of IL-2 in the supernatant was measured by ELISA.

NF- κ B activation assay

NF- κ B activation assays were carried out as described¹.

Western blot analysis and immunoprecipitation

Cell lysis and blotting were carried out as described²⁹. Membranes were blotted with antibodies to Rip2 (Cayman), I κ B, JNK, p38, ERK1/2 (phosphorylated and unphosphorylated forms for each; Cell signaling), IRAK-1, TRAF6 (both Santa Cruz), and MyD88 (StressGen). We carried out immunoprecipitation as described²⁹ using an anti-Flag monoclonal antibody (Sigma).

Northern blot analysis

Macrophages derived from bone marrow were stimulated with 10 ng ml^{-1} LPS for the indicated periods. Preparation of total RNA samples and northern blot analysis were performed as described previously²⁹.

T-cell stimulation for NF- κ B activation

Purified splenic T cells were incubated with anti-CD3 antibodies at the indicated concentration for 30 min on ice. After washing, cells were incubated with anti-hamster immunoglobulin- γ antibody ($100 \mu\text{g ml}^{-1}$; Vector) at 37°C for the indicated periods. Cytoplasmic and nuclear extracts were used for western blot analysis and gel mobility shift assay using a [³²P]dCTP-labelled probe specific for NF- κ B (5'-GGAGTTGAGGGGACT TTCCAGGC-3').

Received 22 November 2001; accepted 24 January 2002.

- Inohara, N., del Peso, L., Koseki, T., Chen, S. & Nunez, G. RICK, a novel protein kinase containing a caspase recruitment domain, interacts with CLARP and regulates CD95-mediated apoptosis. *J. Biol. Chem.* **273**, 12296–12300 (1998).
- McCarthy, J. V., Ni, J. & Dixit, V. M. RIP2 is a novel NF- κ B-activating and cell death-inducing kinase. *J. Biol. Chem.* **273**, 16968–16975 (1998).
- Thome, M. *et al.* Identification of CARDIAK, a RIP-like kinase that associates with caspase-1. *Curr. Biol.* **8**, 885–888 (1998).
- Medzhitov, R. & Janeway, C. Jr. Innate immune recognition: mechanisms and pathways. *Immunol. Rev.* **173**, 89–97 (2000).
- Kelliher, M. A. *et al.* The death domain kinase RIP mediates the TNF-induced NF- κ B signal. *Immunity* **8**, 297–303 (1998).
- Poltorak, A. *et al.* Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* **282**, 2085–2088 (1998).
- Takeuchi, O. *et al.* Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. *Immunity* **11**, 443–451 (1999).
- Hemmi, H. *et al.* A Toll-like receptor recognizes bacterial DNA. *Nature* **408**, 740–745 (2000).
- Alexopoulos, L., Holt, A. C., Medzhitov, R. & Flavell, R. A. Recognition of double-stranded RNA and activation of NF- κ B by Toll-like receptor 3. *Nature* **413**, 732–738 (2001).
- Demuth, A., Goebel, W., Beuscher, H. U. & Kuhn, M. Differential regulation of cytokine and cytokine receptor mRNA expression upon infection of bone marrow-derived macrophages with *Listeria monocytogenes*. *Infect. Immun.* **64**, 3475–3483 (1996).
- Humke, E. W., Shriver, S. K., Starovansnik, M. A., Fairbrother, W. J. & Dixit, V. M. ICEBERG: a novel inhibitor of interleukin-1 β generation. *Cell* **103**, 99–111 (2000).
- Medzhitov, R. *et al.* MyD88 is an adaptor protein in the hToll/IL-1 receptor family signaling pathways. *Mol. Cell* **2**, 253–258 (1998).
- Kawai, T., Adachi, O., Ogawa, T., Takeda, K. & Akira, S. Unresponsiveness of MyD88-deficient mice to endotoxin. *Immunity* **11**, 115–122 (1999).
- Bertin, J. *et al.* Human CARD4 protein is a novel CED-4/Apaf-1 cell death family member that activates NF- κ B. *J. Biol. Chem.* **274**, 12955–12958 (1999).
- Inohara, N. *et al.* Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor- κ B. *J. Biol. Chem.* **274**, 14560–14567 (1999).
- Ogura, Y. *et al.* Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF- κ B. *J. Biol. Chem.* **276**, 4812–4818 (2001).
- Inohara, N., Ogura, Y., Chen, F. E., Muto, A. & Nunez, G. Human nod1 confers responsiveness to bacterial lipopolysaccharides. *J. Biol. Chem.* **276**, 2551–2554 (2001).
- Ogura, Y. *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**, 603–606 (2001).
- Girardin, S. E. *et al.* CARD4/Nod1 mediates NF- κ B and JNK activation by invasive *Shigella flexneri*. *EMBO Rep.* **2**, 736–742 (2001).
- Grownay, J. D. & Dietrich, W. F. High-resolution genetic and physical map of the Lgn1 interval in C57BL/6J implicates Naip2 or Naip5 in *Legionella pneumophila* pathogenesis. *Genome Res.* **10**, 1158–1171 (2000).
- Hugot, J. P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603 (2001).
- Miceli-Richard, C. *et al.* CARD15 mutations in Blau syndrome. *Nature Genet.* **29**, 19–20 (2001).
- Yang, J., Zhu, H., Murphy, T. L., Ouyang, W. & Murphy, K. M. IL-18-stimulated GADD45- β required in cytokine-induced, but not TCR-induced, IFN- γ production. *Nature Immunol.* **2**, 157–164 (2001).
- Leitenberg, D. & Bottomly, K. Regulation of naive T cell differentiation by varying the potency of TCR signal transduction. *Semin. Immunol.* **11**, 283–292 (1999).
- Ruland, J. *et al.* Bcl10 is a positive regulator of antigen receptor-induced activation of NF- κ B and neural tube closure. *Cell* **104**, 33–42 (2001).
- Sun, Z. *et al.* PKC- θ is required for TCR-induced NF- κ B activation in mature but not immature T lymphocytes. *Nature* **404**, 402–407 (2000).

27. Aronica, M. A. *et al.* Preferential role for NF- κ B/Rel signaling in the type 1 but not type 2 T cell-dependent immune response *in vivo*. *J. Immunol.* **163**, 5116–5124 (1999).
28. Celada, A., Gray, P. W., Rinderknecht, E. & Schreiber, R. D. Evidence for a γ -interferon receptor that regulates macrophage tumoricidal activity. *J. Exp. Med.* **160**, 55–74 (1984).
29. Kobayashi, K., Hatano, M., Otaki, M., Ogasawara, T. & Tokuhisa, T. Expression of a murine homologue of the inhibitor of apoptosis protein is related to cell proliferation. *Proc. Natl Acad. Sci. USA* **96**, 1457–1462 (1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank C. L. Stewart and M. Schnare for providing reagents; L. Evangelisti for technical assistance; and F. Manzo and G. Chenell for manuscript preparation. This work was supported by grants from the National Institutes of Health (G.N., J.G., C.A.J., R.M. and R.A.F.) and from the Howard Hughes Medical Institute (C.A.J., R.M. and R.A.F.). K.K. is a recipient of a Postdoctoral Fellowship of the Cancer Research Institute; L.D.H. was supported by a fellowship from the Cancer Research Fund of the Damon Runyon-Walter Winchel Foundation; and R.A.F. and C.A.J. are Investigators and R.M. is an Assistant Investigator of the Howard Hughes Medical Institute.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.A.F. (e-mail: richard.flavell@yale.edu).

An amino-acid taste receptor

Greg Nelson*, Jayaram Chandrashekar*, Mark A. Hoon†, Luxin Feng*, Grace Zhao*, Nicholas J. P. Ryba† & Charles S. Zuker*

* Howard Hughes Medical Institute and Departments of Biology and Neurosciences, University of California at San Diego, La Jolla, California, 92093-0649, USA

† National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland 20892, USA

The sense of taste provides animals with valuable information about the nature and quality of food. Mammals can recognize and respond to a diverse repertoire of chemical entities, including sugars, salts, acids and a wide range of toxic substances¹. Several amino acids taste sweet or delicious (umami) to humans, and are attractive to rodents and other animals². This is noteworthy because L-amino acids function as the building blocks of proteins, as biosynthetic precursors of many biologically relevant small molecules, and as metabolic fuel. Thus, having a taste pathway dedicated to their detection probably had significant evolutionary implications. Here we identify and characterize a mammalian amino-acid taste receptor. This receptor, T1R1+3, is a heteromer of the taste-specific T1R1 and T1R3 G-protein-coupled receptors. We demonstrate that T1R1 and T1R3 combine to function as a broadly tuned L-amino-acid sensor responding to most of the 20 standard amino acids, but not to their D-enantiomers or other compounds. We also show that sequence differences in T1R receptors within and between species (human and mouse) can significantly influence the selectivity and specificity of taste responses.

T1Rs and T2Rs are two families of G-protein-coupled receptors (GPCRs) selectively expressed in subsets of taste receptor cells^{3–11}. T2Rs are involved in bitter taste detection^{4,5}, and T1R2 and T1R3 combine to function as a sweet taste receptor⁷. To identify taste receptors involved in amino-acid detection, we used an expression screening strategy similar to that used in the characterization of bitter and sweet taste receptors. Candidate receptors were expressed in human embryonic kidney (HEK) cells containing the $G\alpha_{16}$ – $G\alpha_Z$

and $G\alpha_{15}$ promiscuous G proteins^{12,13}, and assayed for stimulus-evoked changes in intracellular calcium. In this system, receptor activation leads to activation of phospholipase C β (PLC- β) and release of calcium from internal stores, which can be monitored at the single-cell level using calcium-indicator dyes^{5,7,14}.

Because T1R taste receptors are distantly related to GPCRs that recognize the amino acids glutamate¹⁵ (metabotropic glutamate receptors, mGluRs), GABA¹⁶ (γ -aminobutyric acid; GABA-B receptors) and arginine¹⁷ (the R5-24 receptor), we began by testing members of the T1R family. Patterns of T1R expression define at least three distinct cell types: cells co-expressing T1R2 and T1R3 (T1R2+3, a sweet receptor), cells co-expressing T1R1 and T1R3 (T1R1+3) and cells expressing T1R3 alone⁷. First, we assayed responses of the T1R2+3 sweet taste receptor to all 20 standard and various D-amino acids. Several D-amino acids that taste sweet to humans, and are attractive to mice, trigger robust activation of the T1R2+3 sweet taste receptor (Fig. 1a, b). However, none of the tested L-amino acids activate this receptor.

Mouse T1R1 and T1R3 were transfected alone or in combination and tested for stimulation by L-amino acids. Individual receptors showed no responses. In contrast, T1R1 and T1R3 combine to function as a broadly tuned L-amino-acid receptor, with most amino acids that are perceived as sweet (for example, alanine, glutamine, serine, threonine and glycine²) activating T1R1+3 (Fig. 1). The responses are strictly dependent on the combined presence of T1R1 and T1R3, and are highly selective for L-amino acids; D-amino acids and other natural and artificial sweeteners did not activate the T1R1+3 receptor combination. These results substantiate T1R1+3 as a receptor for L-amino acids, and provide a striking example of heteromeric GPCR receptors radically altering their selectivity by a combinatorial arrangement of subunits.

If T1R1+3 functions as a major L-amino acid taste sensor *in vivo*, we might expect its cell-based behaviour to recapitulate some of the physiological properties of the *in vivo* receptor. Nerve recordings in rats have shown that taste responses to L-amino acids are considerably potentiated by purine nucleotides such as inosine monophosphate (IMP)¹⁸. To assay the effect of IMP, HEK cells expressing the T1R1+3 receptor combination were stimulated with amino acids in the presence or absence of IMP. Indeed, T1R1+3 responses to nearly all L-amino acids were dramatically enhanced by low doses of IMP (Figs 1 and 2a); this effect increased over a range of 0.1–10 mM (Fig. 2b). However, IMP alone elicited no response, even at the highest concentration tested in our assays, and it had no effect on responses mediated by T1R2+3 (either to sweeteners or to L- and D-amino acids; data not shown).

T1R1+3 is prominently expressed in fungiform taste buds⁷, which are innervated by chorda tympani fibres. Therefore, we stimulated mouse fungiform papillae at the front of the tongue with various amino acids in the presence or absence of IMP, and recorded tastant-induced spikes from the chorda tympani nerve. As expected, nerve responses to L-amino acids were significantly enhanced by IMP¹⁸ (Fig. 3). However, IMP had no significant effect on responses to D-amino acids or to non-amino-acid stimuli.

Genetic studies of sweet tasting have identified a single principal locus in mice influencing responses to several sweet substances (the *Sac* locus^{19,20}). *Sac* 'taster' mice are about fivefold more sensitive to sucrose, saccharin and other sweeteners than *Sac* non-tasters. *Sac* codes for T1R3^{7–11,21}. There are two amino-acid differences that define taster and non-taster alleles^{7,9,10}. One of these changes, I60T, introduces a potential glycosylation site that was proposed to eliminate receptor function by preventing receptor dimerization¹⁰. This poses a conundrum because responses to L-amino acids are not influenced by the *Sac* locus^{7,22} (and data not shown). Thus, if T1R3 functions as the common partner of the sweet and amino-acid receptors, we reasoned that the T1R3 non-taster allele must selectively affect the T1R2+3 combination.

We examined the effect of the *Sac* non-taster allele on T1R1 and

T1R2 using biochemical and functional assays. First, we investigated receptor heteromerization by co-immunoprecipitating differentially tagged T1R receptors. In essence, HEK cells were co-transfected with taster and non-taster alleles of T1R3 and either haemagglutinin (HA)-tagged T1R1 or T1R2. Receptor complexes were then immunoprecipitated with anti-HA antibodies, and the association with T1R3 assayed with anti-T1R3 antibodies. Our results demonstrated that the non-taster form of T1R3, much like its taster counterpart, assembles into heteromeric receptors with T1R1 and T1R2 (Fig. 4a). This argues against the possibility that the sweet taste deficits of *Sac* non-taster animals result from failure to assemble heteromeric receptors. Second, we examined the functional responses of T1R2+3 (sweet) and T1R1+3 (amino acid) receptors carrying either the taster or non-taster allele of T1R3. The taster and non-taster alleles of T1R3 generate functionally similar receptors when combined with T1R1, but the non-taster form displays significantly impaired responses when combined with T1R2 (Fig. 4b). Thus, responses to L-amino acids are not affected by the *Sac* locus in mice because *Sac* selectively affects the T1R2+3 receptor combination.

The finding that polymorphism in one of the T1R receptor

subunits differentially affects receptor function suggests that other sequence variations in the amino-acid and sweet receptors may significantly influence tastant sensitivity or selectivity. For example, humans can taste a number of artificial sweeteners that rodents cannot (for instance, aspartame, cyclamate and various sweet proteins²³). Rodent and human T1Rs are only about 70% identical⁷. Therefore, we generated heteromeric receptors consisting of human and rodent T1R subunits and assayed for activation by amino acids and artificial sweeteners. Indeed, the presence of human T1R1 or T1R2 greatly altered the sensitivity (Fig. 4c) and the specificity (Fig. 4d) of the amino-acid and sweet taste receptors. Cells expressing human T1R1 are more than an order of magnitude more sensitive to glutamate than to other amino acids, and cells expressing human T1R2 now robustly respond to aspartame, cyclamate and intensely sweet proteins (Fig. 4d and data not shown). Thus, the nature of the unique partner determines whether the receptor complex will function as a sweet receptor or as an amino-acid receptor, and sequence differences in T1Rs between or within species (for example, polymorphisms in *Sac*) can greatly influence taste perception.

In humans, monosodium L-glutamate (MSG) elicits a unique

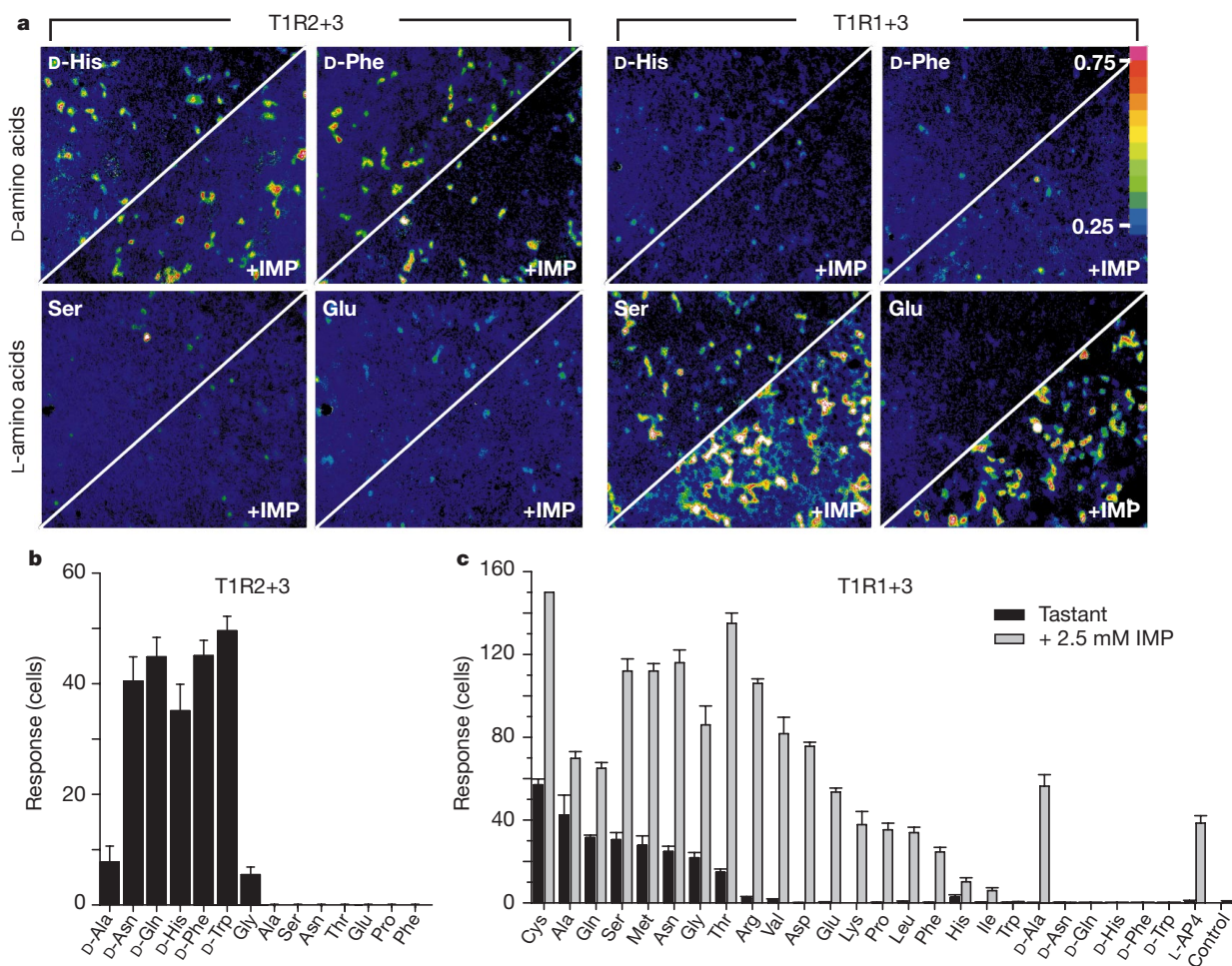


Figure 1 T1R receptor combinations respond differentially to L- and D-amino acids. **a**, HEK-293 cells co-expressing promiscuous G proteins and heteromeric mouse T1R2+3 or T1R1+3 receptors were stimulated with L- and D-amino acids. The T1R2+3 sweet taste receptor is activated by sweet-tasting D-amino acids but not by L-amino acids (left). In contrast, T1R1+3 is activated by L-amino acids and responses are potentiated by IMP (right). Amino acids were 50 mM and IMP was 2.5 mM; the colour scale indicates the F_{340}/F_{380} ratio (see Methods). **b**, **c**, Quantification of amino-acid responses for T1R2+3 (**b**) and T1R1+3 (**c**). Amino acids were 50 mM, and IMP and L-AP4 were 2.5 mM; control

refers to 2.5 mM IMP alone. Each column represents the mean $\pm$ s.e.m. of at least ten independent determinations. IMP had no effect on T1R2+3 (data not shown). D-Amino acids (with the exception of D-Ala in the presence of IMP) and natural or artificial sweeteners did not activate T1R1+3. Trp elicited no responses and Tyr was not assayed because it is insoluble at high concentration. Note that the achiral amino acid Gly activates both receptor complexes. All calcium measurements and quantifications were performed as described in the Methods and ref. 7.

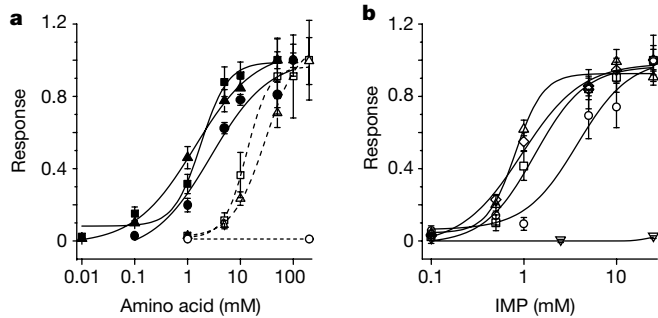


Figure 2 Dose response of T1R1+3 to L-amino acids and IMP. **a**, Dashed lines with open symbols represent dose responses of T1R1+3 with L-amino acids (squares, Ala; circles, Glu; triangles, Ser). The presence of 2.5 mM IMP (solid lines with filled symbols) shifts the responses by at least one order of magnitude to the left. Equivalent results were obtained with most L-amino acids (see also Fig. 1b). **b**, IMP potentiates responses of T1R1+3. Shown are dose responses for Ala (2 mM, squares), Glu (4 mM, circles), Ser (2 mM, triangles), Gly (4 mM, diamonds) and IMP (inverted triangles). Responses were normalized to the mean response at the highest concentration. Each point represents the mean $\pm$ s.e.m. of at least ten assays.

savoury taste sensation called umami^{24,25}. Hallmarks of the umami taste are its potentiation by purine nucleotides, and activation by the mGluR-agonist L-AP4 (ref. 25). Recently, a mGluR4 splice variant has been reported as a candidate umami receptor²⁶. An obvious question is whether T1R1+3 is an umami receptor. Our results demonstrate that T1R1 and T1R3 combine to function as a broadly tuned amino-acid receptor. Notably, T1R1+3 responses to L-AP4 (Fig. 1), MSG and other amino acids are greatly potentiated by purine nucleotides. Thus, we propose that T1R1+3 is a constituent of the umami response. Future studies should help define whether T1R1+3 is the principal, or an additional, umami receptor. An interesting paradox that emerged from this work is the relationship between receptor activity and taste perception. For example, T1R1+3 responds to most L-amino acids, but not all amino acids

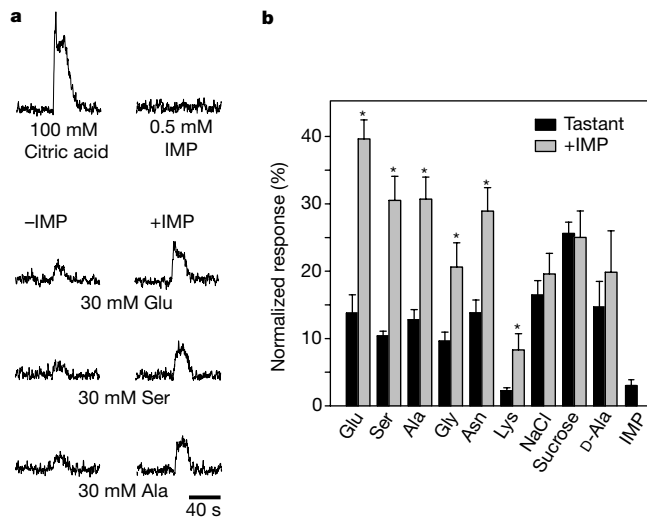


Figure 3 IMP stimulates responses of the chorda tympani nerve to amino acids in mice. **a**, Integrated neural responses of C57BL/6 mice to Glu, Ser and Ala (30 mM each) were recorded with and without 0.5 mM IMP. The responses to 100 mM citric acid and 0.5 mM IMP alone are shown in the upper traces. Equivalent results were obtained for most L-amino acids. **b**, Integrated neural responses, such as those shown in **a**, were normalized to the responses of 100 mM citric acid. Black bars, tastant alone; grey bars, tastant + 0.5 mM IMP. The values are means $\pm$ s.e.m. ($n = 5$). Sucrose was used at 100 mM and all other tastants at 30 mM. Asterisks indicate statistically significant differences ($P < 0.05$).

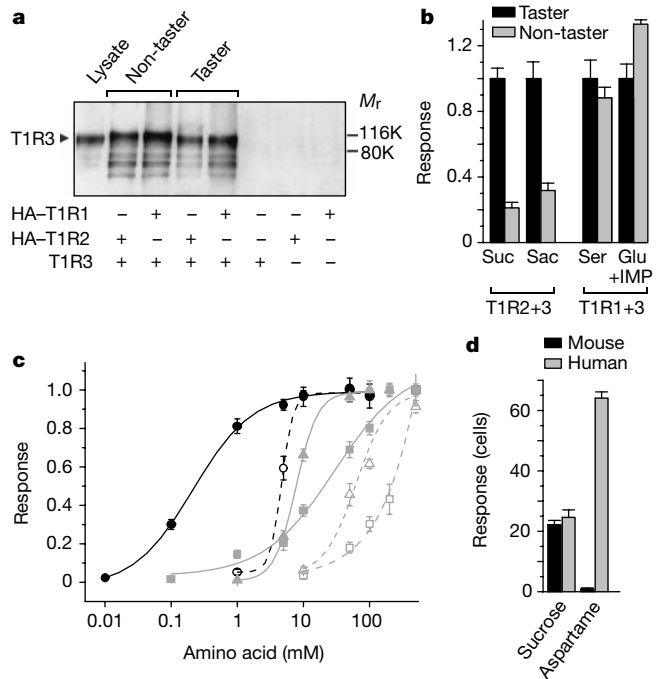


Figure 4 Polymorphic differences in T1Rs influence receptor function.

a, Immunoprecipitation and western blot analyses shows that Sac non-taster and taster alleles of T1R3 form heteromeric complexes with T1R1 and T1R2. Cells were transfected with combinations of T1Rs as indicated. All extracts were immunoprecipitated with anti-HA antibodies, and the resulting protein complexes probed with anti-T1R3 antibodies. M_r , relative molecular mass, in thousands (K). **b–d**, Results of cell-based calcium imaging assays. **b**, The Sac allele selectively affects the T1R2+3 heteromeric receptor. Responses were normalized to the mean responses obtained with the taster allele (black bars). The responses of T1R2+3 to sweet compounds are significantly reduced when the non-taster T1R3 allele is used, but responses of T1R1+3 to amino acids are unaffected, even in the presence of IMP. **c**, Human T1R1 influences sensitivity to monosodium L-glutamate. Low-concentration MSG robustly activates receptors containing human T1R1 (open circles), and IMP potentiates the response (filled circles). Also shown for comparison are dose responses for Ala (squares) and Ser (triangles). For each series, responses were normalized to the mean response at the highest concentration. **d**, Mouse T1R2+3 (black bars) responds to sucrose and other natural and artificial sweeteners, but not aspartame. However, substituting human T1R2 for mouse T1R2 (grey bars) in the rodent T1R2+3 receptor imparts aspartame sensitivity.

taste the same: some are attractive to mice and sweet to humans, whereas others are neutral; some are even perceived as bitter and are aversive to animals². Similarly, very few amino acids elicit the taste of umami. The recent identification of bitter, sweet, and now an amino-acid taste receptor provide a powerful platform to help decode the interplay between the various taste modalities, and the link between events at the periphery (taste receptor cells) and the central nervous system (perception and behaviour). □

Methods

Heterologous expression and calcium imaging

Cells were grown, maintained and transfected exactly as described earlier⁷. Transfection efficiencies were estimated by co-transfection with a green fluorescent protein (GFP) reporter plasmid and were typically $>70\%$. FURA-2 acetomethyl ester was used to measure intracellular calcium concentration ($[Ca^{2+}]_i$), and assay conditions were identical to those previously described⁷. Responses were measured for 60 s and the fluorescence ratio at wavelengths of 340 and 380 nm (F_{340}/F_{380}) was used to measure $[Ca^{2+}]_i$. For data analysis, response refers to the number of cells responding in a field of about 300 transfected cells. Cells were counted as responders if F_{340}/F_{380} increased above 0.27 after addition of tastant. In general, $>90\%$ of the responding cells had $F_{340}/F_{380} > 0.35$. Dose-response functions were fitted using the logistical equation. Studies involving taster and

non-taster alleles of T1R3 used constructs of complementary DNA coding for T1R3 from C57BL/6 and 129/Sv mice, respectively^{7–11,21}.

Immunoprecipitation

Antibodies against T1R3 were generated using a peptide corresponding to residues 824–845 of the mouse receptor. PEAK^{rapid} cells (Edge Biosciences) were transfected with HA–T1R1, HA–T1R2 and T1R3 in various combinations and were gathered and disrupted in buffer containing 50 mM Tris-HCl at pH 7.5, 300 mM NaCl, 1% NP-40, 0.5% w/v sodium deoxycholate, and protease inhibitors (Roche). Lysates were incubated overnight at 4 °C with mouse monoclonal anti-HA antibody (Santa Cruz) and immune complexes were collected with protein A/G–agarose beads. Samples were fractionated by SDS–PAGE, transferred to nitrocellulose membrane and probed with anti-T1R3 antibody. As a control for the specificity of the interactions, we have shown that artificially mixing extracts from cells expressing tagged T1R1 or T1R2 with extracts from cells expressing T1R3 does not produce complexes. Similarly, co-transfection of a Rho-tagged mGluR1 receptor¹⁵ did not produce T1R–GluR1 complexes.

Nerve recording

Lingual stimulation and recording procedures were performed as previously described²⁷. Neural signals were amplified (2,000 ×) with a Grass P511 AC amplifier (Astro-Med), digitized with a Digidata 1200B A/D converter (Axon Instruments), and integrated (r.m.s. voltage) with a time constant of 0.5 s. Taste stimuli were presented at a constant flow rate of 4 ml min^{−1} for 20-s intervals interspersed by 2-min rinses between presentations. All data analyses used the integrated response over a 25-s period immediately after the application of the stimulus. Each experimental series consisted of the application of six tastants bracketed by presentations of 0.1 M citric acid to ensure the stability of the recording. The mean response to 0.1 M citric acid was used to normalize responses to each experimental series.

Received 21 January; accepted 7 February 2002.

Published online 24 February 2002, DOI 10.1038/nature726.

1. Lindemann, B. Taste reception. *Physiol. Rev.* **76**, 718–766 (1996).
2. Iwasaki, K., Kasahara, T. & Sato, M. Gustatory effectiveness of amino acids in mice: behavioral and neurophysiological studies. *Physiol. Behav.* **54**, 531–542 (1985).
3. Hoon, M. A. *et al.* Putative mammalian taste receptors: a class of taste-specific GPCRs with distinct topographic selectivity. *Cell* **96**, 541–551 (1999).
4. Adler, E. *et al.* A novel family of mammalian taste receptors. *Cell* **100**, 693–702 (2000).
5. Chandrashekar, J. *et al.* T2Rs function as bitter taste receptors. *Cell* **100**, 703–711 (2000).
6. Matsunami, H., Montmayeur, J. P. & Buck, L. B. A family of candidate taste receptors in human and mouse. *Nature* **404**, 601–604 (2000).
7. Nelson, G. *et al.* Mammalian sweet taste receptors. *Cell* **106**, 381–390 (2001).
8. Kitagawa, M., Kusakabe, Y., Miura, H., Ninomiya, Y. & Hino, A. Molecular genetic identification of a candidate receptor gene for sweet taste. *Biochem. Biophys. Res. Commun.* **283**, 236–242 (2001).
9. Montmayeur, J. P., Liberles, S. D., Matsunami, H. & Buck, L. B. A candidate taste receptor gene near a sweet taste locus. *Nature Neurosci.* **4**, 492–498 (2001).
10. Max, M. *et al.* Tas1r3, encoding a new candidate taste receptor, is allelic to the sweet responsiveness locus Sac. *Nature Genet.* **28**, 58–63 (2001).
11. Sainz, E., Korley, J. N., Battey, J. F. & Sullivan, S. L. Identification of a novel member of the T1R family of putative taste receptors. *J. Neurochem.* **77**, 896–903 (2001).
12. Offermanns, S. & Simon, M. I. Gα₁₅ and Gα₁₆ couple a wide variety of receptors to phospholipase C. *J. Biol. Chem.* **270**, 15175–15180 (1995).
13. Mody, S. M., Ho, M. K., Joshi, S. A. & Wong, Y. H. Incorporation of Gα₂-specific sequence at the carboxyl terminus increases the promiscuity of Gα₁₆ toward G_i-coupled receptors. *Mol. Pharmacol.* **57**, 13–23 (2000).
14. Tsien, R. Y., Rink, T. J. & Poenie, M. Measurement of cytosolic free Ca²⁺ in individual small cells using fluorescence microscopy with dual excitation wavelengths. *Cell Calcium* **6**, 145–157 (1985).
15. Nakanishi, S. Molecular diversity of glutamate receptors and implications for brain function. *Science* **258**, 597–603 (1992).
16. Kaupmann, K. *et al.* Expression cloning of GABA_B receptors uncovers similarity to metabotropic glutamate receptors. *Nature* **386**, 239–246 (1997).
17. Specia, D. J. *et al.* Functional identification of a goldfish odorant receptor. *Neuron* **23**, 487–498 (1999).
18. Yoshii, K., Yokouchi, C. & Kurihara, K. Synergistic effects of 5'-nucleotides on rat taste responses to various amino acids. *Brain Res.* **367**, 45–51 (1986).
19. Fuller, J. L. Single-locus control of saccharin preference in mice. *J. Hered.* **65**, 33–36 (1974).
20. Lush, I. E. The genetics of tasting in mice. VI. Saccharin, acesulfame, dulcin and sucrose. *Genet. Res.* **53**, 95–99 (1989).
21. Bachmanov, A. A. *et al.* Positional cloning of the mouse saccharin preference (Sac) locus. *Chem. Senses* **26**, 925–933 (2001).
22. Bachmanov, A. A., Tordoff, M. G. & Beauchamp, G. K. Intake of umami-tasting solutions by mice: a genetic analysis. *J. Nutr.* **130**, 935S–941S (2000).
23. Bachmanov, A. A., Tordoff, M. G. & Beauchamp, G. K. Sweetener preference of C57BL/6ByJ and 129P3/J mice. *Chem. Senses* **26**, 905–913 (2001).

24. Ikeda, K. On a new seasoning. *J. Tokyo Chem. Soc.* **30**, 820–836 (1909).
25. Kurihara, K. & Kashiwayanagi, M. Introductory remarks on umami taste. *Ann. NY Acad. Sci.* **855**, 393–397 (1998).
26. Chaudhari, N., Landin, A. M. & Roper, S. D. A metabotropic glutamate receptor variant functions as a taste receptor. *Nature Neurosci.* **3**, 113–119 (2000).
27. Dahl, M., Erickson, R. P. & Simon, S. A. Neural responses to bitter compounds in rats. *Brain Res.* **756**, 22–34 (1997).

Acknowledgements

We thank W. Guo for help generating various T1R expression constructs, and A. Becker for help with tissue culture and antibodies. We are grateful to S. Simon and R. Erikson for teaching us the nerve-recording preparation. We also thank L. Stryer and members of the Zuker lab for help and advice. This work was supported in part by a grant from the National Institute on Deafness and Other Communication Disorders to C.S.Z. C.S.Z. is an investigator of the Howard Hughes Medical Institute.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.S.Z. (e-mail: charles@flyeye.ucsd.edu).

addendum

Virus-mediated killing of cells that lack p53 activity

Kenneth Raj, Phyllis Ogston & Peter Beard

Nature **412**, 914–917 (2001).

Some background information to our work on adeno-associated virus (AAV)-induced apoptosis in cells lacking p53 activity was omitted owing to space constraints. The oncosuppressive activity of parvoviruses has been reviewed^{1,2}. AAV inhibits cell cycle progression³, even when ultraviolet-inactivated⁴, as do AAV-coded Rep proteins⁵. p53-dependent cytopathic effects of parvovirus H1 have been reported⁶. H1 is an autonomous virus that can replicate in cells and lyse them. This is different from AAV, which is defective and does not replicate in the conditions we used. H1 and AAV share little sequence homology and the structures of the DNA termini are not the same. □

1. Rommelaere, J. & Cornelis, J. J. Antineoplastic activity of parvoviruses. *J. Virol. Methods* **33**, 233–251 (1991).
2. Schlehofer, J. The tumor suppressive properties of adeno-associated viruses. *Mutat. Res.* **305**, 303–313 (1994).
3. Hermanns, J. *et al.* Infection of primary cells by adeno-associated virus type 2 results in a modulation of cell cycle-regulating proteins. *J. Virol.* **71**, 6020–6027 (1997).
4. Winocour, E., Callahan, M. & Huberman, E. Perturbation of the cell cycle by adeno-associated virus. *Virology* **167**, 393–399 (1988).
5. Saudan, P., Vlach, J. & Beard, P. Inhibition of S-phase progression by adeno-associated virus Rep78 protein is mediated by hypophosphorylated pRb. *EMBO J.* **19**, 4351–4361 (2000).
6. Telerman, A. *et al.* A model for tumor suppression using H-1 parvovirus. *Proc. Natl Acad. Sci. USA* **90**, 8702–8706 (1993).

Spectroscopy and fluorometry

New ideas include kits — QC Paks — for validating microplate fluorometers.

M400 spectrophotometer

Camspec www.camspec.co.uk

Sensitive attachments

Camspec offers two new accessories for adjusting the sensitivity of its M400 fibre optic diode array UV–visible spectrophotometer. A one-metre path length flow cell increases sensitivity 100-fold by allowing light to pass down a capillary tube of sample as a continuation of the optical fibre. It can be used for discrete or continuous flow sampling. Alternatively, an ATR probe is available to decrease the sensitivity of the spectrophotometer by a factor of 10,000, facilitating absorbance measurements *in situ* on concentrated, undiluted samples. A fibre-optic lead with a length of up to 20 metres and an optional foot switch enable users to perform quality assurance and quality control measurements directly on the production line, with no need to take samples back to the laboratory.

Eagle II XXL

EDAX www.edax.com

Bird's eye view

EDAX's Eagle II XXL X-ray fluorescence elemental analyser has an overall vacuum chamber size of 650 × 650 mm, allowing high-sensitivity X-ray fluorescence measurements to be made on larger samples. It comes in two versions, one with a vertical chamber height of 400 mm for analysis of flat panels, the other with a vertical chamber height of 700 mm for larger, three-dimensional objects. The unit is equipped with a motorized XYZ stage with nominal travel of 300 × 300 × 100 mm. Small, high-intensity X-ray beams are produced by collimators or mono-capillaries in

combination with a 40-W X-ray tube. Applications include forensics, high-tech materials, failure analysis/condition monitoring, semiconductor lab and layer/plating analysis.

CHEMUSB2-VIS spectrophotometer

Ocean Optics www.oceanoptics.com

Small footprint, big impression

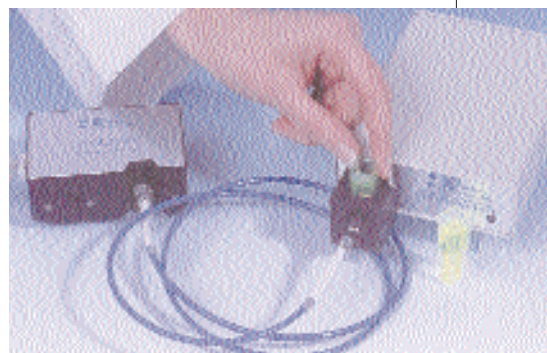
At only 3.5 inches wide and 4 inches long, this device combines a built-in light source, a full diode array spectrometer and a holder for 1-cm cuvettes with high-speed electronic technology to provide real-time, PC-based spectroscopy in the field or in the lab. The instrument captures full spectra from 390 to 900 nm in 25 milliseconds. It can be attached to the USB port of a desktop or notebook PC running Windows 98 or 2000 operating systems. The software automatically loads calibration coefficients and starts data retrieval.

Biowave spectrophotometer

WPA www.wpald.co.uk

Spectrophotometer for molecular biologists

The Biowave spectrophotometer comes pre-programmed with all the standard methods for protein and nucleic acid analysis. The device can accept ultra-micro cuvettes for sample volumes as small as 70 µl. The use of diode array detection technology eliminates the need for a cover over the cuvette chamber. The entire range (200–820 nm) can be scanned quickly and displayed on a large graphic screen. Bidirectional RS232 communication and software for overlaying and manipulating scans are included. A brochure is available from the WPA website, download time 50 seconds at 56 K.



CHEMUSB2-VIS: PC-based spectroscopy.

Xcalibur plug-in

LC Packings BV www.lcpackings.com

Single-point control of UltiMate system

A new upgrade for the Xcalibur plug-in provides full control of the UltiMate Capillary and Nano LC system. With the new plug-in, the UltiMate micro-pump and detection module, including the FAMOS micro-autosampler and the Switchos micro-column switching module, can be controlled using ThermoFinnigan's Xcalibur MS software. The plug-in features high-resolution UV data display, acquisition and reporting at up to four user-selectable wavelengths. LC Packings is a Dionex company (www.dionex.com).

RL785 and RL830 lasers

Renishaw www.renishaw.com

High-power near-IR excitation lasers

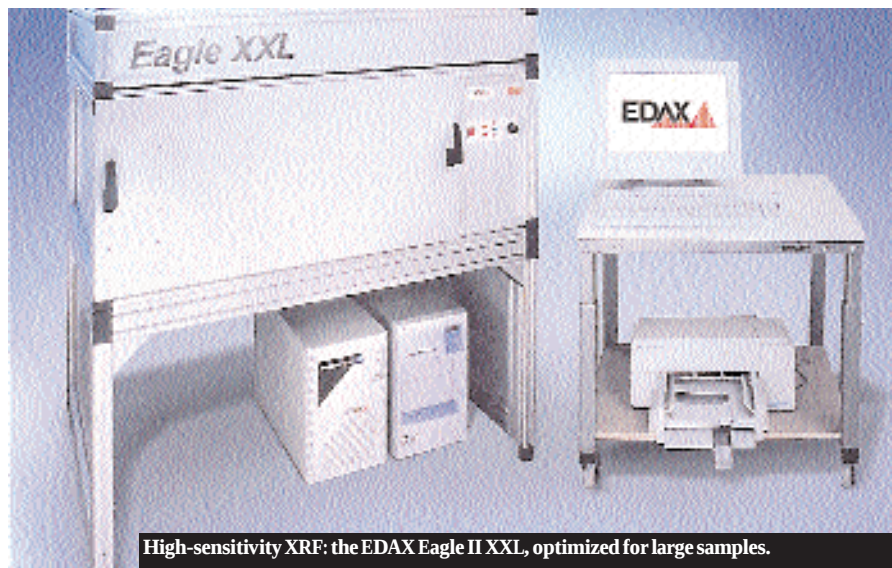
Operating at 785 and 830 nm, Renishaw's compact, low-weight, air-cooled lasers can be used in both portable and OEM laser-based systems. Features include frequency stabilization, side-band suppression and an audible warning of laser emission. Both models provide a power output of ≥ 300 mW at the laser head, with stability of ± 1% when held at ± 2°C within the range 10–35 °C over 16 hours. Options for fibre-optic coupling are available, as is a version suitable for Raman spectroscopy.

FlashMaster Personal

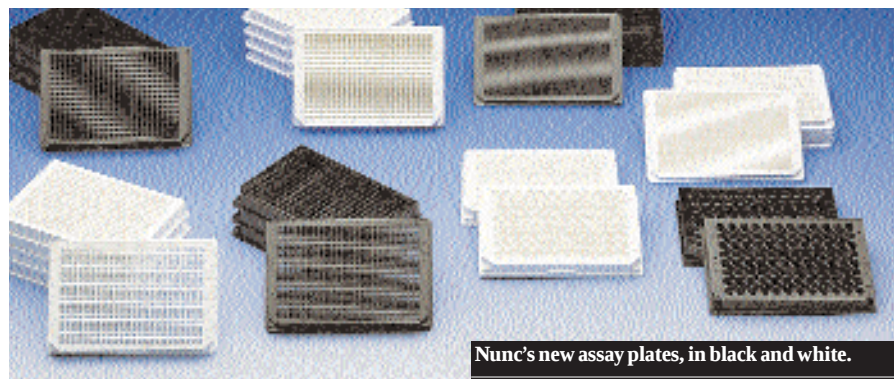
Jones Chromatography www.jones-chrom.co.uk

Chromatography for singles

This addition to the FlashMaster line of flash chromatography purification systems is billed as a low-cost, entry-level, single-column, single-user unit. It offers easy solvent changes, pre-packed columns, unlimited solvent volumes and a wide range of sample sizes. FlashMaster Personal is designed for synthetic chemistry, particularly in drug discovery and medicinal chemistry.



High-sensitivity XRF: the EDAX Eagle II XXL, optimized for large samples.



Nunc's new assay plates, in black and white.

Black and white plates

Nunc A/S www.nuncbrand.com
On the surface, they're all different

Nunc's black and white plates are designed for fluorescence- and luminescence-based assays, respectively. Cell Culture treated plates can be used in cell-based assays, viable cell counts, cell adhesion studies, cytotoxicology and drug sensitivity studies. Maxi-Sorp plates for immunological applications such as ELISAs are optimized for IgG binding. PolySorp plates are used for binding more lipophilic molecules to a solid phase for subsequent ELISA. Non-treated plates for liquid-phase immunoassays or molecular biology complete the range. All plates conform to the SBS footprint and come in 96- 384- and 1,536-well formats, and are available as solid plates or with optical bottoms for flexible reading.

eRecordManager 2001

Thermo LabSystems www.thermolabsystems.com
Future-proof data storage

Spectral and chromatographic data from multiple data formats can be stored using this new software release. Oracle-based security provides authentication auditing, version control and electronic signatures. For long-term archival, eRecordManager automates the translation of over 150 analytical data file formats into a platform-neutral format based on XML. Archive policies can be configured by the user and specific fields extracted from archived data and used for subsequent searching. Data can be organized according to experiments and incorporated into lab reports. The program can be used with the Thermo Galactic suite of desktop spectroscopy software and is Windows-compatible.

Spectrum Spotlight 300

PerkinElmer www.perkinelmer.com
IR imaging in a flash

PerkinElmer uses Duet detector technology in this imaging system, which the company says overcomes the limitations of traditional FT-IR systems by eliminating the need for step-scan data collection. Operating over a wider wavelength range, the Duet detector

generates images with complete chemical information in minutes. Integrated software enables users to identify small differences in chemical content to provide information about defects, impurities and composition. The system can be used to complement existing imaging techniques, such as scanning electron microscopy.

AvaMouse

Avantes www.avantes.com
Handheld reflection spectrometer

AvaMouse consists of an integrated xenon flash lamp that can produce 35 flashes per second and a miniaturized optical bench with a photodiode detector. The handheld device can be connected to a computer's Rs-232 interface and used for reflection measurements in the visible range to determine colour of flat surfaces. Integrated SpectraWin-Color software enables users to make colorimetric measurements to determine L, a, b, hue and chromaticity. A database with multiple reference colours can be built so that dE values can be followed in time-dependent series.

XF138-2/ XF140-2 ALPHA Vivid

Glen Spectra www.jyhoriba.co.uk
Filter-out sample auto-fluorescence

Glen Spectra's Vivid filter sets are designed for far-red microscopy dye systems optimized for various light sources. Fluorescent labelling in the far-red region of the light spectrum is expanding as long-wavelength

dyes nearly eliminate sample auto-fluorescence. The XF138-2 ALPHA Vivid set combines an ALPHA shortpass excitation filter with a 690 nm ALPHA longpass filter and is suitable for use with Ar-Kr mixed gas, He-Ne and red diode laser lines. The XF140-2 ALPHA Vivid set has a wide bandpass excitation filter and can be used with broader band mercury Hg arc excitation energy.

EZ-Label

Pierce www.piercenet.com
Convenient fluorescent labelling

Pierce's new labelling kits contain everything necessary to label large or small proteins, including those in which the amount of protein sample available is small. Each kit includes fluorescent dye in individual microtubes, dimethylformamide, phosphate and borate buffers, pre-packed desalting columns, Slide-A-Lyzer mini dialysis units and amber reaction tubes for benchtop fluorescent labelling. The kits are available in fluorescein, rhodamine and fluorescein isothiocyanate formats and contain the reagents for five fluorescent labelling reactions using up to 10 mg ml⁻¹ protein each.

QC Paks

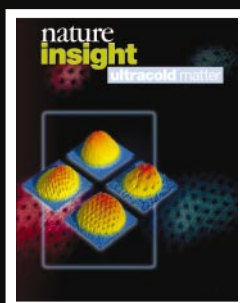
Innovative Instruments www.3i-usa.com
Testing tool for microplate readers

The QC Pak kit can be used to validate the performance of a fluorescence microplate reader. It consists of a microplate with a series of organic dyes embedded in a rigid styrene matrix and an Excel spreadsheet to monitor parameters such as reader precision, linearity, filter integrity, cross-talk between wells and plate alignment. The microplate contains a variety of dyes to assess a broad wavelength range. A fluorescein-like fluorophore is used for statistical validation of the integrity of filters in the blue and red wavelength range. To check linearity, eight samples are analysed and each sample measured five times using the reader's control software.

These notes are compiled in the Nature office from information provided by the manufacturers.

nature insight

Ultracold matter



Cover illustration

Quantum vortices in a rotating condensate of sodium atoms. The images represent two-dimensional cuts through the density distribution and show density minima attributable to the vortex cores. (Courtesy of J. R. Abo-Shaeer, MIT.)

Under unimaginably cold conditions, new types of physical behaviour await our discovery. As the thermal motions of atoms and molecules become progressively smaller, the elusive quantum world — usually masked at higher temperatures — should come into focus.

Superfluidity of liquid helium, first observed in the 1930s, is one of the earliest and most tangible examples. Although this phenomenon occurs at just 2 K, modern laser-cooling technology can attain much lower temperatures. So what happens when we skim absolute zero?

For atomic gases, this question was answered dramatically in 1995 with the production of a quantum state of matter known as a Bose–Einstein condensate. Formed at nanokelvin temperatures, a condensate is a cloud of atoms that pays no heed to everyday physical rules; atomic identities are lost as the particles blur into a coherent quantum entity.

Condensates and superfluids are inextricably linked, yet many questions surround their relationship. The main difference between atomic gas condensates and superfluid helium lies in the strength of the particle interactions, which are much weaker in the former system. This simplifying feature of Bose–Einstein condensates makes them extremely attractive for theoretical studies.

A source of coherent atoms also provides a wonderful practical tool. Just as the development of the laser revolutionized optics, so the ability to generate coherent matter waves opens exciting research possibilities. Nonlinear optical processes such as four-wave mixing can now be carried out using atoms in place of photons; matter-wave amplification has been observed, and atomic analogues of optical ‘squeezed’ states realized.

Although Bose–Einstein condensation is one of the triumphs of laser-cooling technology, it is by no means the only success story. Individual ultracold atoms and ions are also enormously versatile, making an impact in fields such as metrology and quantum information processing.

This collection of review articles covers some of the highlights of recent research into ultracold matter — a field that has seen two Nobel prizes within the past five years.

Karen Southwell Senior Editor

overview:

- 206 Cold atoms and quantum control**
S. Chu

review articles:

- 211 Bose–Einstein condensation of atomic gases**
J. R. Anglin & W. Ketterle
- 219 Nonlinear and quantum atom optics**
S. L. Rolston & W. D. Phillips
- 225 Quantum encounters of the cold kind**
K. Burnett, P. S. Julienne, P. D. Lett, E. Tiesinga & C. J. Williams
- 233 Optical frequency metrology**
Th. Udem, R. Holzwarth & T. W. Hänsch
- 238 Quantum information processing with atoms and photons**
C. Monroe

Editor, *Nature*: Philip Campbell
Insights Editor: Karl Ziemelis
Editorial Assistant: Simon Gibson
Production Editor: Simon Gribbin

Art Director: Majo Xeridat
Diagrams: Ann Thomson
Vicky Askew
Suzanne Coleman

Production Manager: Yvonne Strong
Sponsorship: Emma Jones

Cold atoms and quantum control

Steven Chu

Physics Department, Stanford University, Stanford, California 94305-4060, USA

This overview prefaces a collection of Insight review articles on the physics and applications of laser-cooled atoms. I will cast this work into a historical perspective in which laser cooling and trapping is seen as one of several research directions aimed at controlling the internal and external degrees of freedom of atoms and molecules.

On 10 December 2001, the most spectacular application of laser cooling and atom trapping was recognized on the 100th anniversary of the Nobel prize. The physics prize was awarded to Eric Cornell, Carl Wieman and Wolfgang Ketterle for the “achievement of Bose–Einstein condensation of alkali atoms and for the early fundamental studies of the properties of the condensates”. The demonstration of Bose–Einstein condensation marked the fulfilment of a 70-year-old dream that began with a remarkable prediction of Albert Einstein, building on the calculations of Satyendra Nath Bose. They showed that a gas of atoms, when cooled to sufficiently low temperatures and high densities, would collapse into a single quantum state in which all of the physical properties used to describe each atom, such as its position and velocity, would be the same.

The achievement of Bose–Einstein condensation is one of the highlights of our ability to manipulate the position and velocity of atoms. This technology has led to a flurry of research activity where the pace of fundamental discoveries shows no signs of diminishing. A sample of recent scientific progress is reviewed in this collection of Insight articles, beginning with a review of Bose–Einstein condensation by Anglin and Ketterle on pages 211–218. The work discussed in the following pages can be viewed as an extension of our ability to control the quantum degrees of freedom of atoms. The control of these variables is linked to our ability to place atoms and photons into a well-defined set of quantum states and to manipulate these states coherently.

Control of the internal degrees of freedom of atoms

Radio-frequency techniques

The first step in the coherent control the quantum states of atoms and molecules began with I. I. Rabi, who introduced radio-frequency resonance techniques to molecular beams in 1938 (ref. 1). These methods formed the basis of precision molecular, atomic and nuclear spectroscopy and have served as a standard in the coherent control of atoms.

The outline of a typical apparatus used in the 1940s for atomic-beam resonance is shown in Fig. 1. It consists of a state-selection region, a radio-frequency resonance region and an analyser region. Radio-frequency radiation in the C region is used to excite an atom in an initial state $|a\rangle$ to a final state $|b\rangle$ by tuning the frequency to the atomic resonance, $\omega_{ab} = (E_a - E_b)/\hbar$, where E_a and E_b are the energies of the quantum states $|a\rangle$ and $|b\rangle$. In modern optical experiments, a laser replaces the radio-frequency source, optical pumping methods² are now the preferred method of preparing the initial internal state of the atoms, and fluorescence detection is used instead of the analysing B magnet.

It required four decades before laser methods could match the coherent phase control of the early radio-frequency measurements. There were three primary reasons why the optical control of atomic states lagged behind the radio-frequency and microwave work. First, radio- and microwave-frequency sources were far superior to visible-light sources. Second, the coherent interaction of radio and electric dipole microwave radiation with atoms is not plagued by the dephasing effects of spontaneous emission present in optical transitions. Third, radio-frequency transitions do not have complications arising from Doppler effects.

Radio-frequency radiation was relatively easy to generate because the essential components used to produce an oscillating circuit were made out of parts that were macroscopic in size. After Maxwell's prediction that visible light and heat were oscillating electric and magnetic fields that obeyed a wave equation, Heinrich Hertz turned to radio waves. It was the ability to generate and detect low-frequency radio waves with wave properties such as interference and refraction that made the dramatic connection between visible light and the time-varying fields studied by Michael Faraday.

Atomic clocks and the extension to the optical domain

In 1949, Norman Ramsey extended Rabi's atomic-beam resonance method by dividing the resonance region into two separated, oscillatory-field regions (ref. 3; and see Fig 2). In the first microwave cavity, atoms are excited into an equal superposition of ground and excited states. The atoms then travel through a radiation-free region with their magnetic moments rotating at the precession frequency $\omega_{ab} = (E_a - E_b)/\hbar$, and enter into a second excitation region. If the radiation in the second cavity is tuned to the atomic resonance and is exactly in phase with the radiation in the first cavity, the excitation is completed. After many ($\sim 10^6$) cycles of oscillation between the two zones of interaction, if the radiation is half a cycle out of phase with the precessing atom, the atom is returned to the ground state. This measurement, based on the quantum interference between two internal atomic states, was the first atom interferometer.

The frequency linewidth $\Delta\omega = \Delta E/\hbar$ of the atomic resonance in this type of microwave experiment is limited by the Heisenberg uncertainty principle, $\Delta E \Delta t \geq \hbar$, where Δt is the measurement time. Ramsey's invention allowed the measurement time to be increased by two-to-three orders of magnitude. This measurement technique produces an atomic clock of extraordinary precision and accuracy when the microwave oscillator is locked to the well-defined and reproducible atomic resonance frequency. With the advent of laser cooling, Ramsey's method could be used on an atomic fountain of atoms^{4,5} that increased the measurement time by another two orders of magnitude. Atomic clocks

(Fig. 3) remain one of the most important applications of the coherent interaction of atoms with electromagnetic radiation.

Improvements in our ability to parse time into increasingly self-consistent and reproducible intervals (that is, the 'measurement of time') have been associated with our ability to count the passage of ever-shorter time intervals. In essence, faster oscillators make better clocks. The caesium atomic clock standard, oscillating 9,192,631,770 times a second, has an absolute uncertainty $\Delta\nu/\nu \sim 10^{-15}$. To improve this accuracy by several orders of magnitude, the Ramsey method must be extended to the optical domain.

In the microwave domain, it is straightforward to ensure that all the atoms in an experiment experience the same electromagnetic phase, as the dimensions $\lambda/2$ of the standing wave in the cavities are larger than the transverse spread of the atomic beam. However, if two optical standing waves are used, atoms with different transverse velocities can easily experience trajectories through the apparatus where the overall phase difference between the two Ramsey zones differs by half a cycle, as shown in Fig. 4. Solutions to this problem were found by moving to a geometry with three, separated standing-wave regions^{6,7}, the use of two-photon Doppler-free transitions⁸, or the interaction with four successive running waves of light⁹.

Initially, the atomic recoil due to the light was not considered, but in 1989, Christian Bordé¹⁰ realized that the four-zone geometry^{11,12} created an atom interferometer with spatially separated atomic paths. Thus, the direct extension of Ramsey's separated oscillatory-field method in the microwave regime to the optical domain led to spatially separated atom interferometers. Atom interferometers based on optical transitions (and also methods adapted from work using nuclear magnetic resonance) have led to precise and accurate measurements of gravity¹³, gravity gradients¹⁴, rotations¹⁵ and the photon recoil of an atom¹⁶.

There is one very important distinction between atom interferometers designed for atomic clocks and the atom interferometers developed to measure inertial effects or fundamental constants. Chebotayev, Bordé and their co-workers were motivated to look for a method that would allow them to stabilize a laser to a high-Q optical resonance, whereas the atom interferometers developed by Kasevich and Chu bypassed the requirement of an ultra-stable laser by using Raman transitions between two ground states of the atom¹⁷. The achieved resolution of these atom interferometers ($\sim 10^{-3}$ Hz out of a frequency shift of $\sim 10^7$ Hz) is determined by the frequency difference of the two laser beams used to excite the Raman transition. Because the two laser frequencies can be phase-locked to each other with radio-frequency methods, this configuration allowed complete phase control of the two-photon optical transitions.

Control of the external degrees of freedom of atoms

The development of atomic clocks is one example of the considerable work done in the past 50 years to extend the molecular-beam reso-

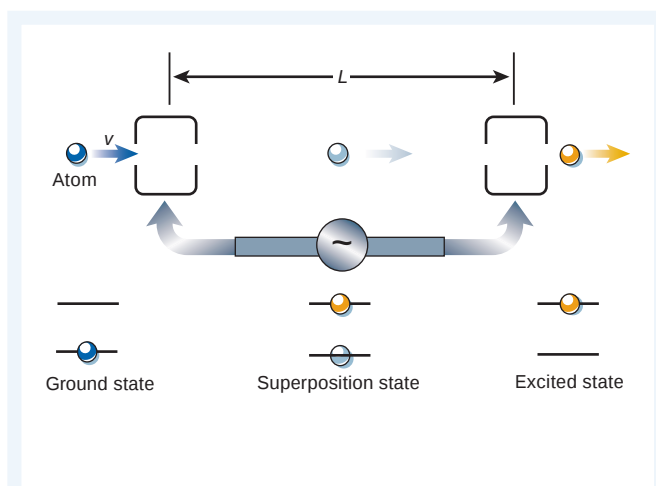


Figure 2 Ramsey's separated oscillatory-field method. The quantum measurement time $\Delta t = L/v$, where v is the velocity of the atom. If the microwave oscillator is tuned to the precise frequency of the atomic resonance, the excitation is completed. As the frequency of the microwave oscillator is varied slightly, the atomic population will oscillate between the ground and excited states, depending on the phase of the microwave radiation relative to the phase difference of the two atomic states.

nance methods and the closely related nuclear magnetic resonance methods into the optical domain. But physical processes exist that do not have counterparts in the radio or microwave region of the electromagnetic spectra, one example being the laser cooling of atoms.

Laser cooling was first demonstrated with ions confined in traps in 1978 (ref. 18). Subsequently, ion-cooling methods were improved to the point where a single ion could be cooled to the lowest vibrational energy level of the trap¹⁹. This capability then led to the creation of quantum states where the electronic and motional degrees of freedom of the ion became intimately connected and could no longer be factored into separate Hilbert spaces²⁰. Much of the early progress in quantum computing (see review in this issue by Monroe, pages 238–246) came from the ability to coherently link and entangle these degrees of freedom. In contrast to ions, the laser cooling of neutral atoms^{21–25} required two more decades to reach the zero-point limit.

Laser cooling requires the transfer of the entropy from the ensemble of atoms being cooled to the radiation field. All the demonstrated cooling methods use spontaneous scattering of radiation as the means of transferring this entropy; unfortunately, however, spontaneous emission also is a heating mechanism. At sufficiently high densities, cooling efficiency degrades owing to inelastic collisions that allow the transfer of energy from the internal to the external degrees of freedom. For that reason, the final stages of cooling used in Bose–Einstein condensation use evaporative cooling^{26,27}.

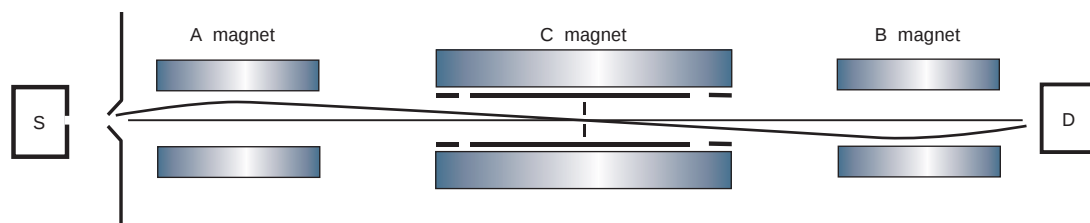
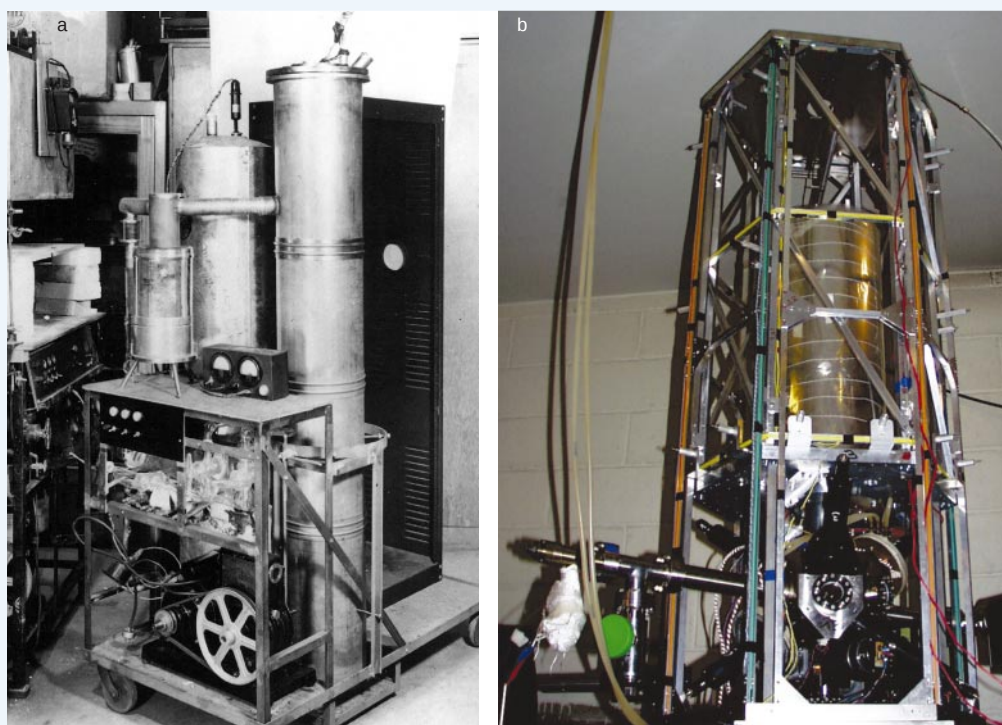


Figure 1 Traditional atomic-beam resonance apparatus. Atoms emerge from a source chamber S into a vacuum chamber and are sent through three regions of magnetic field. The first region consists of an inhomogeneous magnetic field that is used to deflect the atoms with a particular magnetic moment through a slit in the C region. The atoms drift through a C region consisting of a uniform-bias magnetic field and radio

frequency coils, and then through a B region. In the absence of the B magnet, the atoms that pass through the slit will not reach the detector. However, with the proper choice of field gradient, the B magnet will focus atoms in a particular atomic state onto the detector, D.

Figure 3 Atomic clocks. **a**, The first caesium atomic clock, built by Jerrold Zacharias in 1953. Zacharias also proposed a caesium clock using a fountain of atoms to test Einstein's prediction that a clock would slow down in raised in a gravitational potential. He failed to construct an atomic fountain, and although he never published his proposal, his attempt was passed down to two generations of physicists by word of mouth. (Photo: MIT Museum.) **b**, A rubidium atomic fountain clock built by the BNM-LPTF (Bureau National de Métrologie — Laboratoire Primaire du Temps et des Fréquences) and the Ecole Normale Supérieure, Paris. The atoms are trapped and cooled in the lower part of the apparatus before being launched upwards to a height of ~ 1 metre. The estimated accuracy of this clock is 1 part in 10^{15} , or 7 minutes over a period of time equal to the age of the Universe (~14 billion years). (Photo: C. Solomon, ENS.)



Improved laser-cooling methods use spontaneous scattering in the initial stages of cooling, but once a particular atom is cooled, the interaction of that atom with light is suppressed. This elegant trick was introduced in the first laser sideband cooling scheme by Wineland and Dehmelt²⁸, and then repeated in many subsequent cooling schemes, such as velocity-selective coherent population trapping of neutral atoms²⁹, the Raman cooling of free atoms³⁰, Raman sideband cooling of atom in optical lattices³¹, as well as in improved optical traps such as the 'dark' magneto-optic trap³².

Control of spontaneous emission

Remarkably, the spontaneous emission properties of atoms can be altered significantly by placing them in an electromagnetic cavity³³. The electromagnetic fluctuations of the quantum vacuum still have to obey Maxwell's equations, and if a high-*Q* microwave cavity is made to be out of resonance with radiation at an atomic transition, the quantum fluctuations that drive spontaneous emission are greatly reduced³⁴. The suppression of spontaneous emission, first demonstrated with Rydberg atoms in 1983 (refs 35,36), was later extended to the optical regime with an ultra-high-finesse Fabry–Perot cavity that enhanced the probability of spontaneous emission into one mode of the cavity above the probability of spontaneous emission into all other competing vacuum modes³⁷.

The use of electromagnetic cavities is increasingly important in the coherent manipulation of both the internal and external states of atoms and their coupling to the radiation field. For example, a non-resonant cavity (but still one strongly coupled to the atomic transition) was used to mediate a resonant exchange collision. Here, an atom in state $|e_1\rangle$ and another atom in state $|g_2\rangle$ exchange two virtual photons to create an entangled state $|\Psi\rangle = \alpha|e_1, g_2\rangle + \beta|g_1, e_2\rangle$ (ref. 38). There are also several quantum computing proposals based on atomic and photonic quantum states entangled with an optical cavity, discussed by Monroe on pages 238–246.

A number of authors have discussed how an optical cavity can be used to laser cool atoms^{39,40}. For example, Vuletić^{41,42} has proposed cooling atoms with a non-resonant, two-photon scattering process by embedding atoms in three sets of counter-propagating laser beams tuned to the red wavelength side of a Fabry–Perot cavity fringe. Cooling is accomplished by causing the atoms to preferential-

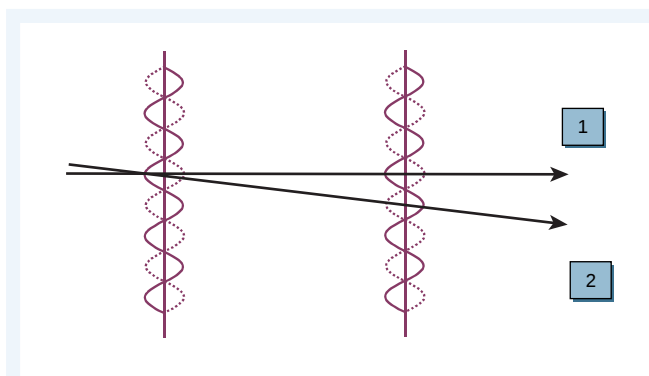


Figure 4 Two atomic trajectories are shown interacting with optical beams of light. In the transit time between the two Ramsey zones, atoms accumulate many cycles of optical phase. Atoms that move along trajectory 1 see a half-cycle phase difference in the number of accumulated cycles of phase relative to the atoms that move along trajectory 2. An average over the spread of transverse velocities in an atomic beam washes out the oscillations in atomic population that would have occurred if all the atoms had the same transverse velocity.

ly scatter blue-shifted light into a resonant cavity mode. Because this cooling process is determined by the cavity resonance, it applies to atoms or molecules with an arbitrary internal level structure.

Control of optical frequencies

The control and measurement of optical sources of light are finally beginning to match and even exceed radio and microwave devices. The stability of laser sources is now starting to surpass the best microwave sources, and broadly tunable, optical-frequency synthesizers with a frequency resolution of one part in 10^{14} will soon be available commercially (T. W. Hänsch, personal communication; and see <http://www.menlosystems.com> for more information on this development). Given the remarkable developments that have occurred within the past few years, the frequency accuracy of an optical clock may reach an uncertainty of one part in 10^{18} within the next decade.

It is worth noting that this revolution in metrology benefited from a number of technological advances. First, the short-term stability of lasers is vastly improved by locking them to an ultra-stable Fabry–Perot cavity with electro- and/or acousto-optic devices⁴³. Second, the high-Q reference cavities are fashioned out of mirrors with scattering and absorption losses of a few parts per million. Third, the long-term stability is supplied by ions confined in traps or neutral atoms in an atomic fountain. Fourth, we now have an elegant method to count directly the optical cycles of phase. This achievement, described in the review in this issue by Udem, Holzwarth and Hänsch on pages 233–237, launches us onto a new epoch in optical metrology by supplying the last critical component needed to make an absolute frequency measurement of light.

Control of collisions

Atomic collisions are generally thought to be messy, incoherent interactions, but the apparent incoherence is the result of our inability to keep track of the large number of degrees of freedom accessible in a typical study of collision. As pointed out Julianne and colleagues (see review in this issue on pages 225–232), all collisions are inherently coherent, as the scattering matrix is unitary.

When atoms are cooled to very low temperatures and large de Broglie wavelengths, atomic scattering is dominated by the lowest allowed angular momentum partial wave: *s* waves for bosons and *p* waves for spin-polarized fermions. With the number of collisional degrees of freedom severely reduced, it has become possible to understand the quantum mechanical collision effects with unmatched precision. For example, photo-association spectroscopy⁴⁴ and threshold resonance spectroscopy based on magnetically tunable (Feshbach) resonances^{45–47} allow the measurement of collision parameters with an accuracy of up to five significant digits. In the case of caesium atoms, the agreement between theory and experiment is remarkable^{48,49}.

Because single ions and atoms can now be confined to the lowest lying quantum state of micro-traps, the spatial extent of the atoms becomes well defined and the duration and strength of collision interactions become controllable. A number of authors have pointed out the power of controlled cold collisions for creating entangled states for use in quantum information processing and spectroscopy^{50–52}. In the near future, the creation of single-mode atom waveguides and the control of the number of interacting atoms through collision blockade effects or Fermi statistics may also be demonstrated. We are entering an era where (sometimes maligned) collisions will enable us to manipulate the internal and external degrees of freedom of atoms. Indeed, coherent, nonlinear collision interactions may become one of the most powerful tools for the creation of deeply entangled quantum states.

Control of atomic de Broglie waves

Atoms in a Bose–Einstein condensate can be described by the same wavefunction and the same phase, and when released out of the trap, they retain memory of their initial phase. With the proper coherent manipulation of the atoms, nonlinear interference experiments such as four-wave mixing and the creation of solitons have now been extended to atomic de Broglie waves^{53–55}. These experiments are examples of the new field of nonlinear and quantum atom optics, which is covered in the review by Rolston and Phillips on pages 219–224 of this issue.

Atoms, however, are not like photons: they have mass and many more internal degrees of freedom. Consequently, nonlinear atom optics will not merely mirror phenomena already seen in the optical domain. Additionally, unlike photons that must interact with each other through a polarizable medium, atoms have strong, direct nonlinear interactions with each other. These interactions have already been used to create non-classical quantum states with number fluctuations well below the shot noise ‘limit’⁵⁶.

One lure of nonlinear atom interactions is the possibility of creating massively entangled states of particles (or to establish non-local

quantum correlations between particles) in two arms of an interferometer. Virtually all photon or atom interferometer experiments carried out so far are based on single-particle interference profoundly enunciated by Dirac 70 years ago: “Each photon interferes only with itself. Interference between two different photons never occurs.”⁵⁷ As a result, the phase uncertainty of interferometers scales as $1/\sqrt{n}$, where *n* is the number of particles contributing to the interference. There are currently a number of proposals to generate deeply entangled, non-classical states⁵⁰ that may allow one to operate an interferometer in the Heisenberg limit where the phase noise scales as $1/n$ (refs 58,59).

Control of many-body and macroscopic wavefunctions

The attractiveness of a Bose–Einstein condensate of a dilute gas is that we have unparalleled control of its properties, together with powerful methods to measure those properties. Long-standing predictions in the weakly interacting domain can be tested for the first time, but that is not where the real excitement lies. Because of the great variation in atom–atom interaction strengths of alkali atoms, and because the *s*-wave scattering lengths can be tuned with a magnetic field from large positive to large negative values, experiments have been done⁶⁰ that previously could not have been contemplated. With suitable choice of scattering length, we can bridge the gap between a nearly non-interacting dilute gas and a strongly interacting quantum fluid.

More generally, these systems may allow us to study the transition between quantum properties of isolated atoms and the rich set of phenomena seen in many-body quantum systems. As an excellent example, a Bose–Einstein condensate embedded in an optical lattice has allowed us to observe a Mott insulator-to-superfluid quantum phase transition⁶¹ in a clean system⁶². In a many-body system at zero temperature, this is characterized by a transition from an ordered solid with integer lattice occupancy to a superfluid, as the quantum coupling between lattice sites is increased relative to the potential well depth (for further details, see review in this issue by Anglin and Ketterle, pages 211–218). Future work may include the study of the effects of static disorder imposed by a stable speckle pattern (Anderson localization) and time varying disorder on this phase transition.

From the perspective of control, the phase transitions provided to us by nature are wonderfully opportune. Bose–Einstein condensation allows us to create large samples of well-ordered atoms long before the atoms are forced into a ground quantum state (where $k_B T$ is much less than the energy-level separation between the ground and first excited states of the trapped atoms) by brute-force cooling. Similarly, the phase transition of the Mott insulator presents us with an entire lattice of (maximally squeezed) atom number states.

With Bose–Einstein condensates, degenerate Fermi gases⁶³ or an array of atom number states as a starting point, exciting applications are inevitable. Not surprisingly, these new research opportunities have stimulated many researchers to think about how to exercise further quantum control over still larger ensembles of atoms and photons, and how to exploit these systems in new ways. To quote Yogi Berra, the noted American philosopher and former catcher for the New York Yankees, “it is difficult to make predictions, especially about the future”. Nevertheless, I predict that the most exciting developments are yet to come. □

1. Rabi, I. I., Zacharias, J. R., Millman, S. & Kusch, P. A new method of measuring nuclear magnetic moments. *Phys. Rev.* **53**, 318 (1938).
2. Kastler, A. in *Nobel Lectures in Physics, 1963–1970* 180–208 (World Scientific, Singapore, 1991).
3. Ramsey, N. F. A new molecular beam resonance method. *Phys. Rev.* **76**, 996 (1949).
4. Kasevich, M. A., Riis, E., Chu, S. & DeVoe, R. G. Rf spectroscopy in an atomic fountain. *Phys. Rev. Lett.* **63**, 612–615 (1989).
5. Laurent, P. et al. in *Laser Spectroscopy XIV* (eds Blatt, R., Escher, J., Liebfried, D. & Schmidt-Kaler, F.) 41–50 (World Scientific, Singapore, 1999).
6. Baklanov, Ye. V., Dubetsky, B. Ya. & Chebotayev, V. P. Non-linear Ramsey resonance in the optical region. *Appl. Phys.* **9**, 171–173 (1976).
7. Chebotayev, V. P., Dubetsky, B. Ya., Kasantsev, A. P. & Yakolev, V. P. Interference of atoms in separated optical fields. *Opt. Soc. Am. B* **2**, 1791–1798 (1985).
8. Beausoleil, R. G. & Hänsch, T. W. Ultra-high resolution two-photon optical Ramsey spectroscopy of

- an atomic fountain. *Phys. Rev. A* **33**, 1661–1670 (1986).
9. Bordé, C. R. On Ramsey fringes in spectroscopy without Doppler broadening. *C.J. Acad. Sci. Ser. B* **284**, 101–104 (1977).
10. Bordé, Ch. Atomic interferometry with internal state labeling. *Phys. Lett. A* **140**, 10–12 (1989).
11. Helmcke, J., Zevgolits, D. & Yen. B. U. Observation of high contrast ultra narrow optical Ramsey fringes in saturated absorption utilizing four interaction zones of traveling waves. *Appl. Phys. B* **28**, 83–84 (1988).
12. Salomon, Ch., Avillier, S., van Lerberghe, A. & Bordé, Ch. in *Laser Spectroscopy VI* (eds Weber, H. P., Lüthy, W.) 159–160 (Springer, Berlin, 1983).
13. Peters, A., Chung, K. Y. & Chu, S. High precision gravity measurements using atom interferometry. *Metrology* **38**, 25–61 (2001).
14. Sadden, M. J., McGrirk, J. M., Boiyyer, P., Hariots, K. G. & Kasevich, M. A. Measurement of the Earth's gravity gradient with an atom interferometer-based gravity gradiometer. *Phys. Rev. Lett.* **81**, 971–974 (1998).
15. Gustavson, T. L., Boyer, P. & Kasevich, M. A. Precision rotation measurements with an atom interferometer gyroscope. *Phys. Rev. Lett.* **78**, 2046–2049 (1997).
16. Hensley, J., Wicht, A., Young, B. & Chu, S. in *Proc. 17th Int. Conf. Atomic Physics* (eds Arimondo, E., De Natale, P. & Inguscio, M.) 43–57 (Am. Inst. Phys., New York, 2000).
17. Kasevich, M. & Chu, S. Atomic interferometry using stimulated Raman transitions. *Phys. Rev. Lett.* **67**, 181 (1991).
18. Wineland, D. J., Drullinger, R. E. & Walls, F. L. Radiation-pressure cooling of bound resonant absorbers. *Phys. Rev. Lett.* **40**, 1639–1642 (1978).
19. Diedrich, F., Bergquist, J. C., Itano, W. M. & Wineland, D. J. Laser cooling to the zero-point energy of motion. *Phys. Rev. Lett.* **62**, 403–407 (1989).
20. Monroe, C., Meekhof, D. M., King, B. E., Itano, W. M. & Wineland, D. J. Demonstration of a universal quantum logic gate. *Phys. Rev. Lett.* **75**, 4714–4717 (1995).
21. Hänsch, T. W. & Schawlow, A. L. Cooling of gases by laser radiation. *Opt. Commun.* **13**, 68–70 (1975).
22. Wineland, D. J. & Itano, I. Laser cooling of atoms. *Phys. Rev. A* **20**, 1521–1540 (1979).
23. Gordon, J. P. & Ashkin, A. Motion of atoms in a radiation trap. *Phys. Rev. A* **21**, 1606–1617 (1980).
24. Letokov, V. S. & Minogin, V. G. Laser radiation pressure on free atoms. *Phys. Rep.* **73**, 1 (1981).
25. Chu, S., Cohen-Tannoudji, C. & Phillips, W. D. 1997 Nobel Lectures in physics. *Rev. Mod. Phys.* **70**, 685–706, 707–720, 721–741 (1998).
26. Hess, H. F. Evaporative cooling of magnetically trapped and compressed spin-polarized hydrogen. *Phys. Rev. B* **34**, 3476–3479 (1986).
27. Mashara, N. *et al.* Evaporative cooling of spin-polarized hydrogen. *Phys. Rev. Lett.* **61**, 935–938 (1988).
28. Wineland, D. & Dehmelt, H. Proposed $10^{14} \Delta\nu < \nu$ laser fluorescence spectroscopy on Ti^+ mono-ion oscillator. *Bull. Am. Phys. Soc.* **20**, 637 (1975).
29. Aspect, A., Arimondo, E., Kaiser, R., Vansteenkiste, N. & Cohen-Tannoudji, C. Laser cooling below the one photon recoil by velocity selective coherent population trapping. *Phys. Rev. Lett.* **61**, 826–829 (1988).
30. Kasevich, M. & Chu, S. Laser cooling below a photon recoil with stimulated Raman transitions. *Phys. Rev. Lett.* **69**, 1741–1743 (1992).
31. Kerman, A. J., Vuletić, V., Chin, C. & Chu, S. Beyond optical molasses: 3D Raman sideband cooling of atomic cesium to high phase space density. *Phys. Rev. Lett.* **84**, 439–442 (2000).
32. Ketterle, W., Davis, K. B., Joffe, M. A., Martin, A. & Pritchard, D. E. High densities of cold atoms in a dark spontaneous force optical trap. *Phys. Rev. Lett.* **70**, 2253–2257 (1993).
33. Purcell, E. M. Spontaneous emission probabilities at radio frequencies. *Phys. Rev.* **69**, 681 (1946).
34. Kleppner, D. Inhibited spontaneous emission. *Phys. Rev. Lett.* **47**, 233–237 (1981).
35. Goy, P., Raimond, J. M., Gross, M. & Haroche, S. Observation of cavity enhanced single atom spontaneous emission. *Phys. Rev. Lett.* **50**, 1903–1907 (1983).
36. Hulet, R. G., Hilfer, E. S. & Kleppner, D. Inhibited spontaneous emission by a Rydberg atom. *Phys. Rev. Lett.* **55**, 2137–2140 (1985).
37. Berman, P. R. (ed.) *Cavity Quantum Electrodynamics* (Academic, San Diego, 1994).
38. Osnaghi, S. *et al.* Coherent control of an atomic collision in a cavity. *Phys. Rev. Lett.* **87**, 037902–1–037902–4 (2001).
39. Mossberg, T. W., Lewenstein, A. & Gauthier, D. J. Trapping and cooling of atoms in a vacuum perturbed in a frequency dependent manner. *Phys. Rev. Lett.* **67**, 1723–1726 (1991).
40. Gangl, M. & Ritsch, H. Cooling neutral particles in multimode cavities without spontaneous emission. *J. Mod. Opt.* **47**, 2741–2753 (2000).
41. Vuletić, V. & Chu, S. Laser cooling of atoms, ions, or molecules by coherent scattering. *Phys. Rev. Lett.* **84**, 3787–3790 (2000).
42. Vuletić, V., Chan, H. W. & Black, A. T. Three dimension cavity Doppler cooling and cavity sideband cooling by coherent scattering. *Phys. Rev. A* **64**, 033405–1–033405–7 (2001).
43. Young, B. C. *et al.* in *Laser Spectroscopy XIV* (eds Blatt, R., Eschner, J., Leibfried, D. & Schmidt-Kaler, F.) 61–70 (World Scientific, Singapore, 1999).
44. Stwalley, W. C. & Wang, H. Photoassociation of ultra-cold atoms: a new spectroscopic technique. *J. Mol. Spect.* **195**, 194–228 (1999).
45. Feshbach, H. A unified theory of nuclear reactions II. *Ann. Phys.* **19**, 287–313 (1962).
46. Tiesinga, E., Verhaar, B. J. & Stoof, H. T. C. Threshold and resonance phenomena in ultracold ground-state collisions. *Phys. Rev. A* **47**, 4144–4122 (1993).
47. Inouye, S. *et al.* Observation of Feshbach resonances in a Bose–Einstein condensate. *Nature* **392**, 151–154 (1998).
48. Chin, C., Vuletić, V., Kerman, A. J. & Chu, S. High resolution Feshbach spectroscopy of cesium. *Phys. Rev. Lett.* **85**, 2717–2720 (2000).
49. Leo, P. J., Williams, C. J. & Julienne, P. S. The collision properties of ultracold ^{133}Cs atoms. *Phys. Rev. Lett.* **85**, 2721–2724 (2000).
50. Jaksch, D., Briegel, H.-J., Cirac, J. I., Gardiner, C. W. & Zoller, P. Entanglement of atoms via cold controlled collisions. *Phys. Rev. Lett.* **82**, 1975–1978 (1999).
51. Sørensen, A. & Mølmer, K. Spin-spin interaction and spin squeezing in an optical lattice. *Phys. Rev. Lett.* **83**, 2274–2277 (1999).
52. Helmer, K. & Li, Y. Creating massive entanglement of Bose-Einstein condensed atoms. *Phys. Rev. Lett.* **87**, 170402–1–170402–4 (2001).
53. Deng, L. *et al.* Four-wave mixing with matter waves. *Nature* **398**, 218–220 (1999).
54. Denschlag, J. *et al.* Generating solitons by phase engineering of a Bose-Einstein condensate. *Science* **287**, 97–100 (2000).
55. Burger, S. *et al.* Dark solitons in Bose-Einstein condensates. *Phys. Rev. Lett.* **83**, 5198–5201 (1999).
56. Orzel, C., Tuchman, A. K., Teselaui, M. L., Yasuda, M. & Kasevich, M. A. Squeezed states in a Bose-Einstein condensate. *Science* **291**, 2386–2389 (2001).
57. Dirac, P. A. M. *The Principles of Quantum Mechanics* 3rd edn, p. 9 (Oxford Univ. Press, Oxford, 1947).
58. Heitler, W. *The Quantum Theory of Radiation* 3rd edn (Oxford Univ. Press, Oxford, 1954).
59. Holland, M. J. & Burnett, K. Interferometric detection of optical phase shifts at the Heisenberg limit. *Phys. Rev. Lett.* **71**, 1355–1358 (1993).
60. Donley, E., Anderson, B. P. & Wieman, C. E. New twists in Bose-Einstein condensation. *Opt. Photon. News* Oct. issue, 34–37 (2001).
61. Fisher, M. P. A., Weichman, P. B., Grinstein, G. & Fisher, D. S. Boson localization and the superfluid-insulator transition. *Phys. Rev. B* **40**, 546–570 (1989).
62. Greiner, M., Mandel, O., Esslinger, T., Hänsch, T. W. & Bloch, I. Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms. *Nature* **415**, 39–44 (2002).
63. DeMarco, B. & Jin, D. S. Onset of Fermi degeneracy in a trapped atomic gas. *Science* **285**, 1703–1706 (1999).

Acknowledgements

This work was supported in part by grants from the National Science Foundation, the Air Force Office of Scientific Research, the Department of the Navy and the National Reconnaissance Office.

Bose–Einstein condensation of atomic gases

James R. Anglin & Wolfgang Ketterle

Research Laboratory for Electronics, MIT–Harvard Center for Ultracold Atoms, and Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

The early experiments on Bose–Einstein condensation in dilute atomic gases accomplished three long-standing goals. First, cooling of neutral atoms into their motional ground state, thus subjecting them to ultimate control, limited only by Heisenberg’s uncertainty relation. Second, creation of a coherent sample of atoms, in which all occupy the same quantum state, and the realization of atom lasers — devices that output coherent matter waves. And third, creation of a gaseous quantum fluid, with properties that are different from the quantum liquids helium-3 and helium-4. The field of Bose–Einstein condensation of atomic gases has continued to progress rapidly, driven by the combination of new experimental techniques and theoretical advances. The family of quantum-degenerate gases has grown, and now includes metastable and fermionic atoms. Condensates have become an ultralow-temperature laboratory for atom optics, collisional physics and many-body physics, encompassing phonons, superfluidity, quantized vortices, Josephson junctions and quantum phase transitions.

The lure of lower temperatures has attracted physicists for much of the past century, and with each advance towards absolute zero, new and rich physics has emerged. Laypeople may wonder why freezing cold is not cold enough; but imagine how many aspects of nature we would miss if we lived on the surface of the Sun. Without inventing refrigerators, we would know only gaseous matter and never observe liquids or solids. Cooling to normal earthly temperatures reveals these dramatically different states of matter, but this is only the beginning: many more states appear with further cooling. The approach into the kelvin range was rewarded with the discovery of superconductivity in 1911 and of superfluidity in ^4He in 1938. Cooling into the millikelvin regime revealed superfluidity of ^3He in 1972. The advent of laser cooling in the 1980s opened up a new approach to ultralow-temperature physics. Microkelvin samples of dilute atom clouds were generated and used for precision measurements and studies of ultracold collisions. Nanokelvin temperatures were necessary to explore quantum-degenerate gases, such as the Bose–Einstein condensates (BECs) first realized in 1995. Each of these achievements in cooling has been a significant advance, and recognized with a Nobel prize.

The essential techniques for making quantum-degenerate gases are cooling techniques, because at high temperatures a dilute gas of atoms behaves classically. As long as the atoms’ de Broglie wavelength $\lambda_{\text{dB}} = \hbar/(2Mk_{\text{B}}T)^{1/2}$ is small compared to the spacing between atoms, one can describe their motion with classical trajectories. (λ_{dB} is the position uncertainty associated with the thermal momentum distribution, and increases with decreasing temperature T and atomic mass M .) Quantum degeneracy begins when λ_{dB} and the interatomic distance become comparable. The atomic wave packets overlap, and the gas starts to become a ‘quantum soup’ of indistinguishable particles. If the atoms are bosons, a condensate — a cloud of atoms all occupying the same quantum state — appears at a precise temperature (which, for an ideal gas, is related to the peak atomic density

n by $n\lambda_{\text{dB}}^3 = 2.612$). If the atoms are fermions (see Box 1), cooling gradually brings the gas closer to being a ‘Fermi sea’ in which exactly one atom occupies each low-energy state.

Creating a BEC or a Fermi sea is thus simple in principle — make a gas extremely cold. In most cases, however, quantum degeneracy would simply be pre-empted by the more familiar transitions to a liquid or solid. This more conventional condensation can be avoided only at extremely low densities, about one-hundred-thousandth the density of normal air, so that the formation time of molecules or clusters by three-body collisions (which is proportional to the square of the inverse density) is stretched to seconds or minutes. Because the rate of binary elastic collisions drops only proportionally to the density, these collisions are much more frequent and let the gas equilibrate within about 10 ms, so that degeneracy can be achieved in an effectively metastable gas phase. However, such ultralow density lowers the temperature requirement for quantum degeneracy into the nanokelvin range.

Sub-microkelvin temperatures are reached by combining two procedures. Laser cooling precools the gas so that it can be confined in a magnetic trap¹. In the second stage — forced evaporative cooling² — the trap depth is reduced, allowing the most energetic atoms to escape while the remaining atoms rethermalize at steadily lower temperatures (see ref. 3 for more details). Most experiments with BECs reach quantum degeneracy between 500 nK and 2 μK , at densities between 10^{14} and 10^{15} cm^{-3} . The largest condensates are of 30 million atoms in Na, and a billion in H; the smallest are just a few hundred atoms. Depending on the magnetic trap, the shape of the condensate is either approximately round, with a diameter of 10–50 μm , or cigar-shaped with a diameter about 15 μm and length 300 μm . The full cooling cycle that produces a condensate may take from a few seconds to as long as several minutes.

This review summarizes the recent progress in and beyond Bose–Einstein condensation. Since 1995 this field has grown explosively, drawing researchers from the communities of atomic physics, quantum optics and condensed matter physics. The trapped ultracold vapour has emerged

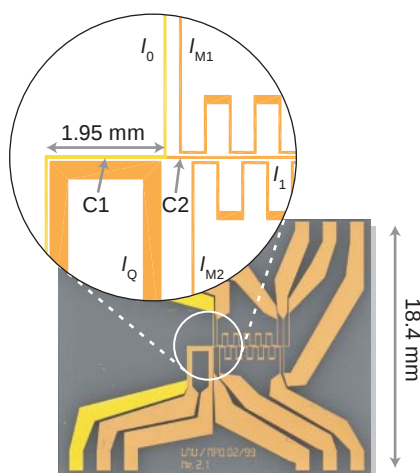


Figure 1 Atom chip. Bose–Einstein condensates (BECs) were created in magnetic traps formed by tiny gold wires that were created lithographically on a substrate^{16,17}. The figure shows the pattern used by researchers at the Ludwig-Maximilians University in Munich¹⁷. Atoms were trapped at positions C1 and C2.

as a new quantum system that is unique in the precision and flexibility with which it can be manipulated. Our field is now at a historic turning point, in which we are moving from studying physics in order to learn about atom cooling to studying cold atoms in order to learn about physics. We begin our review by summarizing new experimental techniques, and then focus on the new ultralow-density condensed matter physics which has been explored.

New techniques and new systems

Evaporative cooling requires a favourable ratio of good to bad collisions — that is, the rate of elastic collisions, which establish thermal equilibrium, must be higher than the rates of inelastic and background gas collisions, which lead to trap loss or molecule formation. A poor collision ratio has hitherto prevented quantum degeneracy being reached in Cs (ref. 4), and until 2000 only ⁸⁷Rb (ref. 5), ²³Na (ref. 6), ⁷Li (ref. 7) and H (ref. 8) had been Bose-condensed, and most of them only in one specific hyperfine state. But since then, research groups have been able to condense Na (A. Görlitz *et al.*, in preparation) and Li (ref. 9) in both upper and lower hyperfine states, and He (refs 10,11), ⁸⁵Rb (ref. 12) and K (ref. 13) have been added to the element list. The condensed He atoms are in an excited (but metastable) electronic state, and their internal energy of 20 eV is released when they strike a surface. This process allows their detection with high efficiency, permitting studies of atomic correlations with single-atom counting resolution in analogy with single-photon counting experiments in optics. One might well have expected this internal energy to be released also in binary collisions, making evaporative cooling impossible. An important early theoretical contribution¹⁴ showed this would not in fact occur, and thus gave experimentalists confidence to proceed.

Further flexibility in condensation is offered by the recent development of ‘all optical’ cooling, in which evaporation is conducted in an optical dipole trap formed of CO₂ laser beams¹⁵, instead of the usual magnetic traps. Reducing the laser intensity allows atoms to escape only at saddle points of the potential, instead of the all-round escape from a magnetic trap during radio-frequency-induced evaporation. But the optical trap holds the gas at a higher density, so that the increased thermalization rate compensates for less efficient evaporation, and Bose–Einstein condensation is quickly reached. Purely optical confinement can be applied to atoms with magnetic moments

too small for magnetic trapping, and also offers a route to condensation for atoms that suffer high spin-flipping losses in magnetic traps.

Magnetic traps have also been advanced. They are now offering new capabilities, through miniaturization. Trapping forces are proportional to magnetic field gradients, so shrinking traps onto microchips, using lithographically created arrays of current-carrying wires to generate fields, produces very tight confinement and also reduces power requirements. Several groups are now developing this technology — two have succeeded in Bose-condensing Rb in this kind of environment (Fig. 1)^{16,17}, while a group at MIT has demonstrated the capability to load a pre-existing condensate into a micro-trap¹⁸. Further progress in these directions may eventually lead to waveguides and beamsplitters for coherent matter, composing microscopic inertial sensors of unprecedented sensitivity. And it will allow study of quantum fluids in restricted geometries.

While sharply varying magnetic fields can confine atoms into waveguides, smooth background fields can also deliver profound control over atoms, by ‘tuning’ their collisional properties. At particular field strengths (Feshbach resonances) the energy of a molecular state may be shifted to zero. This allows two colliding atoms to form a temporary bound state, which causes marked changes in their interaction¹⁹. Such Feshbach resonances were first observed in 1998 (refs 20,21), and have enabled condensation of ⁸⁵Rb (ref. 12).

Another way to compensate for unfavourable collisional properties of an atomic species is to involve another kind of atom, especially one for which evaporative cooling is very effective, as a refrigerant. Such sympathetic cooling was demonstrated between atoms in the two hyperfine states of ⁸⁷Rb (ref. 22), the two Rb isotopes²³, and has enabled the condensation of K by cooling it in collisions with Rb atoms¹³.

Sympathetic cooling is crucial for cooling degenerate fermions, because the Pauli exclusion principle suppresses collisions among fermions of the same species at low temperatures. Significant Fermi degeneracy was first observed through cooling with two hyperfine states of fermionic K (ref. 24), and has recently been obtained by cooling ⁶Li with ⁷Li (refs 9,25) or with Na (ref. 26), or by using two hyperfine states of ⁶Li (ref. 27; and Box 1 and Fig. 2). So far, no experiments have achieved cooling to less than 25% of the Fermi temperature, and further progress is necessary before we can expect to see any pronounced phenomena in fermionic gases, such as Cooper pairing and superfluidity²⁸. The cooling may be limited by Pauli

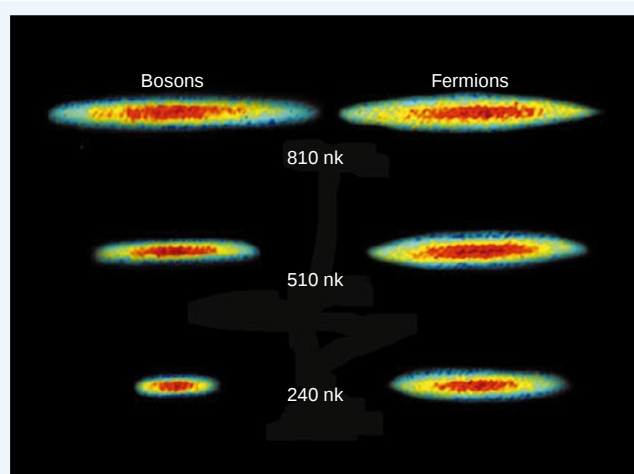


Figure 2 Demonstration of Fermi pressure²⁵. The size of the atom clouds in the magnetic trap shrinks as the temperature is reduced by evaporative cooling. Comparison between bosonic ⁷Li (left) and fermionic ⁶Li (right) shows the distinctive signature of quantum statistics. The fermionic cloud cannot shrink below a certain size determined by the Pauli exclusion principle. This is the same phenomenon that prevents white dwarf and neutron stars from shrinking into black holes. At the highest temperature, the length of the clouds was about 0.5 mm.

blocking of collisions among fermions, by suppression of collisions as a result of superfluidity when the refrigerant is a BEC²⁹, or simply by heating³⁰. Experimental efforts have just started to investigate these issues.

The family of quantum-degenerate gases is growing rapidly. In addition to optical traps, Feshbach resonances and sympathetic cooling, new techniques such as buffer-gas cooling with cryogenic He gas³¹ or cavity cooling^{32,33} may extend the ultracold realm to more kinds of atoms, and even to molecules. There is also the prospect of using photoassociation to make molecular condensates from atomic ones: when two atoms collide they can be stimulated into a long-lived molecular state by applying laser beams³⁴. There is even speculation that such photoassociative 'superchemistry'³⁵ might allow coherent cycling of a system between an atomic and a molecular condensate. It is not only the addition of new species, but also progress towards more complex and more refined experiments that now allows the investigation of a remarkable range of physical phenomena.

Ultralow-density condensed matter physics

A condensate is an ultralow-density condensed matter system. Although it is a gas 100,000 times thinner than air, its temperature is so low that even the weak interactions between atoms create effects typical of a 'conventional' condensed system, such as phase transitions, phonons, superfluidity and Josephson oscillations. With multi-component condensates, and confinement in optical lattices, a wide vista opens. The condensed matter physics of ultracold gases is only beginning.

Atoms interact

The first experiments on BECs showed that they were not ideal gases. When more and more atoms were added to a condensate, it swelled beyond the size of the ground state of the trap — a clear sign of repulsive interactions between the atoms³⁶ — or it collapsed owing to attractive interactions between the atoms³⁷. Without these interactions, the BEC would be an ideal gas with properties similar to the photons in the optical laser. The interactions make the BEC a rich, many-body system that displays phenomena such as sound and superfluidity. An attractive feature of Bose–Einstein condensation in

Box 1

White dwarfs in the laboratory

When physicists speak in general terms of 'particles', they are actually grouping together two very different kinds of objects, whose distinction becomes manifest at low temperatures. It is possible for arbitrarily many particles in the boson class to occupy a single quantum state, like infinitely gregarious hotel guests that are always willing to crowd into a single room. Particles in the fermion class insist on single occupancy.

Extreme cold effectively reduces the size of the hotel, so that the antisocial nature of fermions becomes important. This manifests itself in Fermi degeneracy pressure, by which a cold cloud of fermionic gas resists being compressed into a smaller volume. This effect keeps white dwarf and neutron stars from collapsing into black holes, and has also recently been observed for ultracold gases in the laboratory (ref. 25; and Fig. 2). Unlike the abrupt phase transition of Bose condensation, the onset of Fermi degeneracy is gradual as temperature drops.

At even lower temperatures, if the effective interaction between fermions is attractive, fermions can bind into pairs, the pairs behaving as bosons. Because electrons and ³He atoms are fermions, this effect is responsible for the condensate-like behaviour of superconductors and superfluid ³He. Producing a paired superfluid in ultracold fermionic vapor is a challenge currently being pursued by several groups.

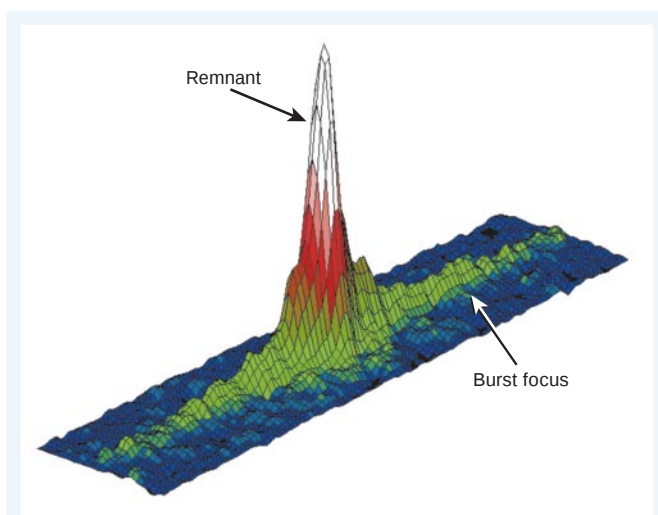


Figure 3 Explosion of a condensate with attractive interactions⁵⁵. By tuning external magnetic fields close to a Feshbach resonance, the interaction between atoms was suddenly switched from zero to attractive. During the collapse, which bears some analogies with a supernova, some atoms were ejected in jets. Because of the magnetic trap, the burst atoms came to a radial focus and were imaged (image size, 60 μm $\times$ 310 μm).

dilute atomic gases is that it can be described theoretically from first principles. Therefore, condensates have become a valuable testing ground for the study of interacting many-body systems³⁸.

The basic theory of the weakly interacting Bose gas was developed from the late 1940s to the early 1960s, and requires that binary collisions are much more frequent than three-body collisions. This condition is fulfilled when the separation between atoms $n^{-1/3}$ is much larger than the s-wave scattering length a , that is, the quantity $na^3 \ll 1$ (typically, $na^3 \approx 10^{-6}$). The magnitude of the scattering length gives the effective range of the interatomic forces (typically 1–5 nm for alkali atoms). The stability of large condensates requires repulsive interactions (positive a). For attractive interactions (negative a), the condensate becomes unstable against collapse if it grows above a certain size.

Early theoretical work has led to the Gross–Pitaevskii equation^{39,40}, which is a wave equation for the macroscopic matter-wave field, and to Bogoliubov's theory of quantum fluctuations around the coherent field⁴¹. In almost all current experiments, the weakly interacting condition is well fulfilled, and the Gross–Pitaevskii–Bogoliubov theory describes the observed phenomena well. But it has received some modern refinements. The original theories violated the conservation of atoms, and several authors have developed number-conserving formulations^{42,43}. The behaviour of condensates at finite temperatures is a frontier of many-body physics^{44,45}, the experimental exploration of which has so far concentrated mainly on the initial formation of condensates.

Condensate growth

The growth of a condensate is an interesting dynamical process — atoms must find the lowest energy state of the system, and long-range coherence has to be established. Experimentally, this process is observed after fast evaporative cooling, which cools the gas below the transition temperature for Bose–Einstein condensation, but is faster than the growth of the condensate to its equilibrium size^{46–48}. A full theoretical description must include the condensate and its elementary excitations, and the interactions with the cloud of thermal atoms (those not part of the condensate). This quantum kinetics problem has been approached from the perspective of quantum optics, which models the condensation process after lasing^{49–51}, while condensed matter theorists have investigated the same process in terms of symmetry breaking and phase relaxation^{52,53}.

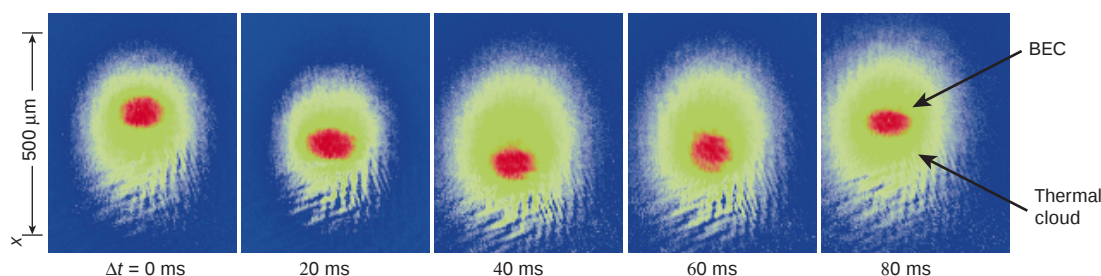


Figure 4 Signature of superfluidity in a Bose-condensed cloud⁶³. The condensate and its thermal ‘halo’ of normal gas respond differently when they are dragged through a periodic potential. In the experiment, the clouds were displaced in a magnetic trap superimposed by an optical lattice. The condensate, distinguished by its much higher

density (colour coded in red), tunneled through the potential peaks and oscillated in the magnetic trapping potential, whereas the normal fraction was pinned by the optical lattice. Interaction between the two clouds eventually led to damping of the condensate motion.

The simplest result is an S-shaped growth curve that reflects initial Bose-stimulated accelerating growth⁴⁹. Observations of the H condensate with its strong two-body losses⁴⁸, and of two timescales in the growth of Rb condensates⁴⁷, add further richness to this non-equilibrium process.

Special dynamics are associated with condensates with attractive interactions. A collapse of the condensate can be triggered by adding atoms to the condensate⁵⁴ or by changing the scattering length through a Feshbach resonance⁵⁵. The observed dynamics of the collapse and subsequent ejection of particles (Fig. 3) are not yet understood theoretically.

Excitations and sound

A normal gas that is as dilute as experimental condensates is in the ‘collisionless regime’, in which a local perturbation of the density simply diffuses away, unless it extends over a distance longer than the particles’ mean free path. [A harmonic trap is special in the sense that it refocuses density fluctuations and transforms a diffusive mode into oscillatory behaviour (but this behaviour is not related to sound).] But in a condensate, collisions that scatter atoms back into the highly occupied quantum state are enormously enhanced, producing a coherent pressure that can support density waves (‘zero sound’) whose wavelength may be far shorter than the mean free path. Early studies of collective excitations in condensates focused on shape oscillations and their damping. More recently, several nonlinear phenomena have been explored, including coupling between modes⁵⁶, soliton propagation^{57,58} and quantized vortices (see below).

Superfluidity in a dilute gas

Superfluidity is commonly defined as flow without dissipation. Evidence for superfluidity in liquid He was obtained in 1938, and

although superfluidity was almost immediately connected by London to Einstein’s theory of Bose–Einstein condensation, it took decades before experimental evidence of a condensate was established with neutron scattering and quantum evaporation⁵⁹. In contrast, in dilute gases condensation was identified first, but it has taken researchers several years to find ways to reveal aspects of superfluidity in these tiny gas clouds.

One way to identify superfluidity is by its characteristic way of breaking down at a precise critical velocity. Only above such a critical velocity is the kinetic energy of the flow sufficient to create excitations. Landau’s early theory identified this with the speed of zero sound, because in a flow below this speed phonon production would not be energetically possible. For a macroscopic flow, in all known superfluids, the critical velocity is usually smaller owing to the excitation of vortices.

Critical velocities in BECs were first studied by moving a focused laser beam through the condensate (stirring it)^{60,61}, and by ‘sloshing’ the condensate in a corrugated potential⁶². When a magnetic force acted on the condensate and a thermal cloud in such an optical lattice, the condensate moved by coherent tunnelling (Fig. 4), whereas the more energetic thermal cloud was pinned⁶³. This counterintuitive behaviour illustrates one of the mysteries of superfluidity and macroscopic quantum mechanics.

Another way to probe a system for superfluid behaviour without ‘touching’ it is to study torsional modes. If a container of liquid He is

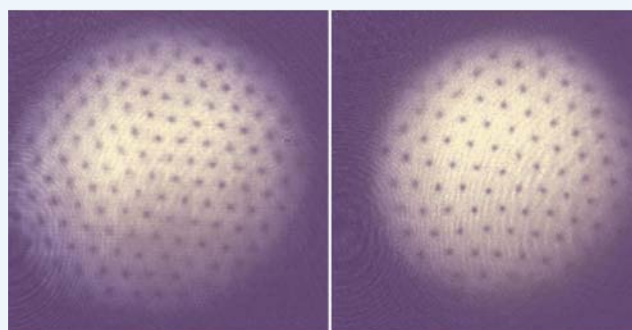


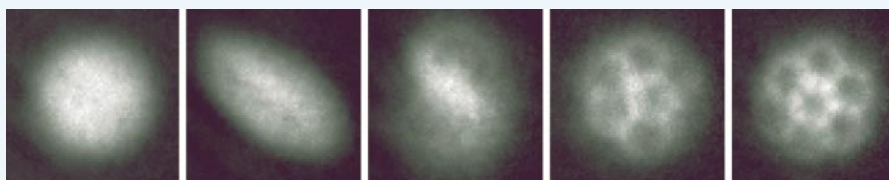
Figure 5 Vortex lattices in rotating BECs⁶⁹. A Na condensate (diameter 60 μm, length 250 μm) was set in rotation by rotating laser beams. It then formed a regular triangular lattice of vortices. Subsequent ballistic expansion resulted in a twenty-times magnification. The images represent two-dimensional cuts through the density distribution and show the density minima due to the vortex cores. The left panel shows a perfect triangular ‘Abrikosov’ lattice, which, on the right side, has a dislocation. The diameter of the clouds was about 1 mm. (Reprinted with permission from ref. 69. Copyright 2001 American Association for the Advancement of Science.)

Box 2

Quantized vortices

Quantum mechanics and the wave nature of matter have subtle manifestations when particles have angular momentum, or more generally, when quantum systems are rotating. When a quantum-mechanical particle moves in a circle, the circumference of the orbit has to be an integer multiple of the de Broglie wavelength. This ‘quantization rule’ leads to the Bohr model and the discrete energy levels of the hydrogen atom. For a rotating superfluid that is described by a macroscopic wavefunction, it leads to the quantization of circulation and quantized vortices. This makes it impossible for a superfluid to rotate as a rigid body — in order to rotate, it must swirl. Vortices are a key feature of superfluid systems.

Figure 6 Vortex generation. One route to vortex formation in a BEC is shown in a series of images taken every 150 ms (ref. 75). A condensate was subjected to a rotating anisotropy for 300 ms. An initially small perturbation grew into a strong quadrupolar deformation and finally into a microscopic ‘tsunami’ that broke into quantized vortices. The field of view in each image is 300 μm . (Copyright 2001 by the American Physical Society.)



slowly rotated, the superfluid fraction does not rotate. Similarly, a condensate held in a yawing trap swings in an irrotational ‘scissors mode’, the frequency of which is inconsistent with rigid body motion^{64,65}. The condition of irrotationality can be violated in a superfluid only by the appearance of quantized vortices (Box 2).

Vortices

It took until 1999 for vortices to be realized experimentally, predominantly because some initial failures made other projects seem more attractive. Following a theoretical proposal⁶⁶, researchers at the University of Colorado at Boulder constructed a quantized vortex in a two-component condensate by imprinting its phase pattern with laser and radio-frequency fields⁶⁷. A few months later, a group at the Ecole Normale Supérieure in Paris used a rotating laser beam to spin up a condensate, and observed vortex arrays⁶⁸. Similar experiments at MIT, with much larger condensates, produced highly regular triangular lattices of vortices (ref. 69; and Fig. 5). Recently, the Boulder group has formed vortices by cooling a rotating normal gas through the transition for Bose–Einstein condensation⁷⁰, and has also managed to create vortex rings⁷¹. These research teams, together with a group at the University of Oxford⁷², have now revealed several aspects of vortex dynamics, including their motion, ‘crystallization’ into lattices and dissipative escape.

It is perhaps the investigation of the threshold for vortex formation, however, that best illustrates the fruitful interplay between theory and experiment in the field of Bose–Einstein condensation. Vortices in liquid He are usually nucleated at surface roughnesses; but Bose condensates are confined in perfectly smooth ‘magnetic containers’, and have been stirred with well characterized rotating potentials. Condensates are therefore an ideal testbed for microscopic theories of vortex generation, which attempt to predict the critical rotational velocity above which vortices become stable. Experimentally, vortices are observed only above a critical rotation frequency, whereas in theory, there may be several velocities that are each critical in different senses⁷³. Which is relevant?

The Paris group found the critical rotation rate close to the quadrupole shape resonance of the condensate (Fig. 6). Instabilities (‘anomalous modes’) of vortices already within the condensate were initially proposed to determine this frequency⁷⁴; but collective dynamical instabilities, associated with the resonance in the vortex-free cloud, were later shown to develop into vortices^{75,76}. Further numerical and analytical results now indicate that a surface mode version of the Landau theory accurately yields the minimum rotation rate for vortex formation^{77,78}, and that above this rate the energetic barrier to vortex penetration is also absent, so that tunnelling is not required⁷⁹. Three-dimensional simulations (numerically integrating the Gross–Pitaevskii equation) have begun to probe this complex behaviour (refs 80,81; and Fig. 7).

Multi-component condensates

New physics emerges when different atomic species are mixed and cooled to quantum degeneracy. In the future, mixtures of bosons may allow studies of interpenetrating superfluids. Mixtures of fermions and bosons, as recently realized experimentally^{9,25,26}, may extend studies of ^3He – ^4He mixtures into new parameter regimes.

Some studies of miscibility, immiscibility and metastability have already been performed using different hyperfine states of Rb or Na

(refs 82,83). New phenomena arise when the different components are converted into each other. Atoms can then be in superposition states, and show spin textures, spin waves and coupling between spin and superfluid flow. In addition to quantized vortices, which like ordinary vortices are line-like structures, there can be bubble-like monopoles^{84,85}. Whereas vortices and monopoles both require a zero-density core, which is a line or a point, respectively, there can exist textures^{86–88} in which the order parameter field twists around in a topologically nontrivial way that does not require a core with vanishing density.

It has also been shown that condensates with spin and antiferromagnetic interactions possess highly correlated singlet ground states⁸⁹, which are quite different from the macroscopic wavefunction of simple condensates. But unfortunately this exotic ground state is vulnerable to small magnetic fields, which favour a symmetry-breaking macroscopic wavefunction.

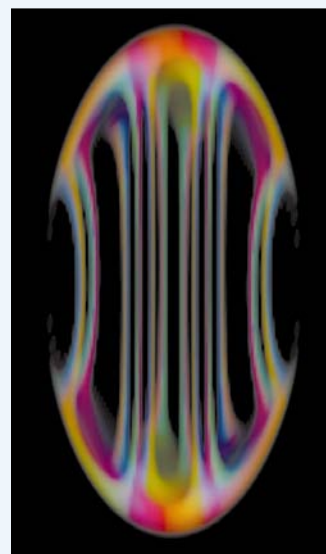
Bose–Einstein condensation in lower dimensions

We have discussed above how the dimensionality of the order parameter space can be raised in multi-component systems. Another opportunity to explore new physics is the reduction of the dimensionality of physical space to two or one dimensions.

To confine a system to lower dimensions usually enhances quantum features and gives rise to new phenomena. A famous example is the quantum Hall effect in a two-dimensional electron gas. Using the tools of atomic physics, condensates can be prepared in a large variety of shapes. Recently, one- and two-dimensional condensates were prepared in highly elongated magnetic and pancake-shaped optical traps⁹⁰.

It is a well known theorem that Bose–Einstein condensation cannot occur in systems that are effectively one-dimensional. Yet the practical implications of this theorem are not that significant for cold-atom experiments, which involve finite sizes and numbers of particles, rather than the thermodynamic limit of textbook theory. For an ideal Bose gas of N atoms in a one-dimensional harmonic trap

Figure 7 Visualization of vortex lines in a trapped condensate. These results were obtained from numerical integration of the Gross–Pitaevskii mean-field theory in three dimensions⁸¹. Shown is the condensate density in inverse-rendering mode, that is, low-density regions are depicted bright, and regions beyond the Thomas–Fermi shell have been discarded. As a result, vortex cores within the condensate appear as bright filaments. Colour indicates the phase of the condensate wavefunction, a phase wrap of 2π yielding the quantized circulation of each vortex. (Image courtesy of D. Feder and P. Ketcham, National Institute of Standards and Technology.)



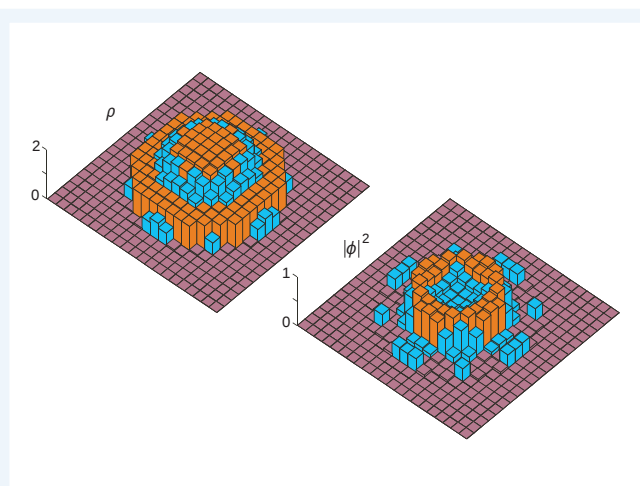


Figure 8 Mott insulator and superfluid phases coexisting in a BEC in a magnetic trap with a superimposed optical lattice⁹⁸. Shown are the total density of cold atoms (left) and the superfluid density (right). Colours are visual guides only. In the insulator phase, the number of atoms shows integer plateaus (left). The superfluid density is maximized between the plateaus. (Copyright 1998 by the American Physical Society.)

with harmonic frequency ω , below a critical temperature $N\omega/\ln N$ there is a steep increase in population of the ground state, which rises smoothly to approach 100% in a manner similar to Bose–Einstein condensation in a three-dimensional trap⁹¹. But the effect of interactions in such effectively one-dimensional systems can be strong. Experiments now support theoretical predictions⁹² that quasi-condensates, with large phase fluctuations, appear at low temperatures⁹³. Another possibility is the realization of a Luttinger liquid⁹⁴. And at extremely low temperatures and densities, wave-mechanical effects make atoms become impenetrable to each other when they propagate in waveguides even if the width of the radial confinement is much larger than the atomic size. This suggests the existence of a transition to a more ordered phase than Bose–Einstein condensation, such as the hard-core gas first analysed by Tonks⁹⁵.

Phase versus number

The ideal or weakly interacting condensate represents a classical matter-wave field in the same way as an optical laser emits a classical electromagnetic wave. As in the optical case, where non-classical light has been widely studied, it is possible to create non-classical states of quantum-degenerate matter. Matter-wave fields can exhibit various kinds of squeezing, as quantum corrections to the classical field theory revise the balance between complementary variables (for example, the density and the phase of coherent matter; see Box 3). Interactions

Box 3

Phase and density fluctuations

Just as light waves possess both intensity and phase, matter waves have density and phase. In quantum-mechanical terms, density and phase are connected by a Heisenberg uncertainty relation in much the same way as position and momentum; they cannot both be precisely defined at once. Condensates produced by current techniques naturally have phase uncertainties on the order of $N^{-1/2}$, where N is the number of atoms in the condensate.

When two water containers are connected, the water level will be the same in both containers, as any other distribution of the water will cost extra energy. Similarly, when two BECs can exchange particles through a barrier, there are conditions where the repulsive energy between atoms will favour an equal distribution of atoms in which the relative density is more sharply defined than in a ‘standard’ condensate. Inhibiting atoms from moving between two parts of a cloud thus blurs any difference in phase between the two regions. It is somewhat counterintuitive, but this blurring of relative phase can be used to make atom interferometers more precise, that is, to measure other phases with an accuracy that is better than the so-called ‘shot noise’ limit.

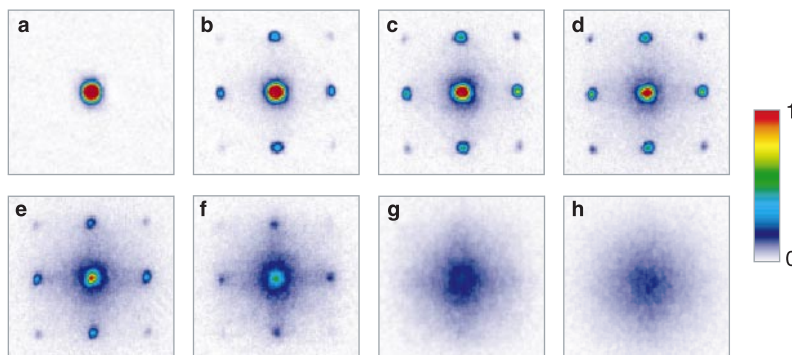
between the atoms can be exploited to create entanglement and squeezing and to ‘engineer’ new non-trivial wavefunctions. This can lead to new forms of quantum matter, and to ensembles that allow higher precision of measurements, for example in atom interferometers⁹⁶.

Tunnelling and Josephson junctions

Phase coherence between two separated condensate samples allows the atomic population to oscillate back and forth between them. Individual atoms shuttle back and forth by quantum tunnelling through the energy barrier that separates the samples — this is the classical Josephson effect, which is well described by the Gross–Pitaevskii equation and represents superfluid flow. When the strength of the repulsive interatomic interactions becomes large compared to the tunnelling rate, then the coherent tunnelling ceases abruptly. The increase in energetic costs for density fluctuations leads to a sharply fixed population difference and an indeterminate relative phase (Box 3); because the Josephson current depends on the phase, it ceases.

The onset of number squeezing has been observed in experiments at Yale University, in which a condensate was subjected to a one-dimensional optical-lattice potential, formed by a standing wave of laser light⁹⁷. Increasing laser power raised the barriers between neighbouring potential minima, suppressing tunnelling

Figure 9 Experimental observation of the quantum phase transition from a superfluid to a Mott insulator in a Rb BEC⁹⁹. When the gas was released from the optical lattice the atoms from the more than 100,000 lattice sites overlapped. In the superfluid state, an interference pattern of multiple peaks formed as a result of the phase coherence. When the lattice potential was increased, strong number squeezing in the insulating state suppressed the interference. From **a** to **g**, the lattice potential was increased from zero to twenty recoil energies.



between them, and the number of atoms trapped in each lattice minimum became more sharply defined. When the optical lattice was suddenly turned off, the few dozen mini-condensates overlapped, but with random relative phases, so that interference effects typical of condensates were suppressed.

Quantum phase transition

A more pronounced version of this effect occurs in a three-dimensional lattice, where a quantum phase transition was predicted theoretically from a dilute superfluid to a Mott insulator (ref. 98; and Fig. 8). In the insulator phase, the lattice sites are occupied by the same small number of atoms, generating a highly ordered state. The transport of atoms from site to site is suppressed by an energy gap that results from atom–atom repulsions. This phase transition has been observed recently by researchers at the Ludwig-Maximilians University in Munich, Germany (ref. 99; and Fig. 9). In these experiments the degenerate gas cloud was allowed to equilibrate in the lattice, and was then released. If it expanded from a superfluid phase, strong interference peaks were created; but above a critical lattice strength these abruptly disappeared, indicating the loss of phase coherence between atoms in neighbouring lattice sites. Experiments with cold atoms are thus beginning to create new condensed matter systems, in which a vast range of phenomena may be realized.

Atomic and molecular physics

In this article, we have emphasized BECs as a new system for condensed matter physics, with novel ways to create many-particle wavefunctions. As discussed by Julienne and colleagues elsewhere in this issue (see pages 225–232), the condensate provides a new laboratory also for collision physics at zero energy — the study of the wavefunctions of two and three particles. Over the past few years, intense experimental and theoretical efforts have elucidated atomic interactions and scattering processes near zero energy¹⁰⁰. Because these interactions depend on the position of individual quantum levels, they can be modified by external magnetic fields through Zeeman shifts¹⁹. Such Feshbach resonances in atomic collisions are now used to produce ‘designer condensates’ with adjustable attractive or repulsive interactions.

Condensation is also important as a superior way to create single-particle wavefunctions (occupied with many identical particles): it enlarges the scope of atom optics by providing ‘atom lasers’. Condensates are atom sources with high brightness and small divergence and are being used for further advances in atom interferometry and other areas of atom optics, as discussed by Rolston and Phillips in their review on pages 219–224 of this issue. When the interactions of the atoms are involved, atom optics becomes nonlinear and processes such as four-wave mixing and soliton propagation occur, so that the underlying physics overlaps with aspects of condensed matter physics. Previously independent physical disciplines are thus drawn together in the study of BECs.

Outlook

The field of quantum-degenerate gases is at an exciting stage of development. Many of the highlights discussed in this paper have been accomplishments of the past year, and are evidence of the field’s continuing rapid development. New atomic systems, in particular ultracold fermions, have considerably broadened the research agenda. Although for bosons the ultimate low-temperature phenomenon — Bose–Einstein condensation — has been accomplished, fermions still pose the challenge of reaching the even lower temperatures that will yield the phase transition into pairing and superfluidity. So the quest for new phenomena at ever lower temperatures will continue. □

1. Arimondo, E., Phillips, W. D. & Strumia, F. *Laser Manipulation of Atoms and Ions* (North-Holland, Amsterdam, 1992).
2. Masuhara, N. *et al.* Evaporative cooling of spin-polarized atomic hydrogen. *Phys. Rev. Lett.* **61**, 935–938 (1988).

3. Ketterle, W., Durfee, D. S. & Stamper-Kurn, D. M. in *Bose-Einstein Condensation in Atomic Gases* (eds Inguscio, M., Stringari, S. & Wieman, C. E.) 67–176 (IOS Press, Amsterdam, 1999).
4. Guéry-Odelin, D., Söding, J., Desbiolles, P. & Dalibard, J. Is Bose-Einstein condensation of atomic cesium possible? *Europhys. Lett.* **44**, 25–30 (1998).
5. Anderson, M. H., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. Observation of Bose-Einstein condensation in a dilute atomic vapor. *Science* **269**, 198–201 (1995).
6. Davis, K. B. *et al.* Bose-Einstein condensation in a gas of sodium atoms. *Phys. Rev. Lett.* **75**, 3969–3973 (1995).
7. Bradley, C. C., Sackett, C. A., Tollet, J. J. & Hulet, R. G. Evidence of Bose-Einstein condensation in an atomic gas with attractive interactions. *Phys. Rev. Lett.* **75**, 1687–1690 (1995).
8. Fried, D. G. *et al.* Bose-Einstein condensation of atomic hydrogen. *Phys. Rev. Lett.* **81**, 3811–3814 (1998).
9. Schreck, F. *et al.* Quasipure Bose-Einstein condensate immersed in a Fermi sea. *Phys. Rev. Lett.* **87**, 080403-1–080403-4 (2001).
10. Robert, A. *et al.* A Bose-Einstein condensate of metastable atoms. *Science* **292**, 461–464 (2001).
11. Pereira Dos Santos, F. *et al.* Bose-Einstein condensation of metastable helium. *Phys. Rev. Lett.* **86**, 3459–3462 (2001).
12. Cornish, S. L., Claussen, N. R., Roberts, J. L., Cornell, E. A. & Wieman, C. E. Stable ⁸⁵Rb Bose-Einstein condensates with widely tunable interactions. *Phys. Rev. Lett.* **85**, 1795–1798 (2000).
13. Modugno, G. *et al.* Bose-Einstein condensation of potassium atoms by sympathetic cooling. *Science* **294**, 1320–1322 (2001).
14. Shlyapnikov, G. V., Walraven, J. T. M., Rahmanov, U. M. & Reynolds, M. W. Decay kinetics and Bose condensation in a gas of spin-polarized triplet helium. *Phys. Rev. Lett.* **73**, 3247–3250 (1994).
15. Barrett, M. D., Sauer, J. A. & Chapman, M. S. All-optical formation of an atomic Bose-Einstein condensate. *Phys. Rev. Lett.* **87**, 010404-1–010404-4 (2001).
16. Ott, H., Fortagh, J., Schlottnerbeck, G., Grossmann, A. & Zimmermann, C. Bose-Einstein condensation in a surface microtrap. *Phys. Rev. Lett.* **87**, 230401-1–230401-4 (2001).
17. Hänsel, W., Hommelhoff, P., Hänsch, T. W. & Reichel, J. Bose-Einstein condensation on a microelectronic chip. *Nature* **413**, 498–501 (2001).
18. Gustavson, T. L. *et al.* Transport of Bose-Einstein condensates with optical tweezers. *Phys. Rev. Lett.* **88**, 020401-1–020401-4 (2002).
19. Tiesinga, E., Verhaar, B. J. & Stoof, H. T. C. Threshold and resonance phenomena in ultracold ground-state collisions. *Phys. Rev. A* **47**, 4114–4122 (1993).
20. Inouye, S. *et al.* Observation of Feshbach resonances in a Bose-Einstein condensate. *Nature* **392**, 151–154 (1998).
21. Courteille, P., Freeland, R. S., Heinzen, D. J., van Abeelen, F. A. & Verhaar, B. J. Observation of a Feshbach resonance in cold atom scattering. *Phys. Rev. Lett.* **81**, 69–72 (1998).
22. Myatt, C. J., Burt, E. A., Ghrist, R. W., Cornell, E. A. & Wieman, C. E. Production of two overlapping Bose-Einstein condensates by sympathetic cooling. *Phys. Rev. Lett.* **78**, 586–589 (1997).
23. Bloch, I., Greiner, M., Hänsch, O. M. W. & Esslinger, T. Sympathetic cooling of ⁸⁵Rb and ⁸⁷Rb. *Phys. Rev. A* **64**, 021402-1–021402-4 (2001).
24. DeMarco, B. & Jin, D. S. Onset of Fermi degeneracy in a trapped atomic gas. *Science* **285**, 1703–1706 (1999).
25. Truscott, A. G., Strecker, K. E., McAlexander, W. I., Partridge, G. B. & Hulet, R. G. Observation of Fermi pressure in a gas of trapped atoms. *Science* **291**, 2570–2572 (2001).
26. Hadzibabic, Z. *et al.* Two species mixture of quantum degenerate Bose and Fermi gases. Preprint cond-mat/0112425 at <http://xxx.lanl.gov> (2001).
27. Granade, S. R., Gehm, M. E., O’Hara, K. M. & Thomas, J. E. Preparation of a degenerate, two-component Fermi gas by evaporation in a single beam optical trap. Preprint cond-mat/0111344 at <http://xxx.lanl.gov> (2001).
28. Stoof, H. T. C. & Houbiers, M. in *Bose-Einstein Condensation in Atomic Gases* (eds Inguscio, M., Stringari, S. & Wieman, C. E.) 537–553 (IOS Press, Amsterdam, 1999).
29. Timmermans, E. & Côté, R. Superfluidity in sympathetic cooling with atomic Bose-Einstein condensates. *Phys. Rev. Lett.* **80**, 3419–3423 (1998).
30. Timmermans, E. Degenerate Fermion gas heating by hole creation. *Phys. Rev. Lett.* **87**, 240403-1–240403-4 (2001).
31. Kim, J. *et al.* Buffer-gas loading and magnetic trapping of atomic europium. *Phys. Rev. Lett.* **78**, 3665–3668 (1997).
32. Horak, P., Hechenblaikner, G., Gheri, K. M., Stecher, H. & Ritsch, H. Cavity-induced atom cooling in the strong coupling regime. *Phys. Rev. Lett.* **79**, 4974–4977 (1997).
33. Vuletic, V. & Chu, S. Laser cooling of atoms, ions, or molecules by coherent scattering. *Phys. Rev. Lett.* **84**, 3787–3790 (2000).
34. Wynar, R., Freeland, R. S., Han, D. J., Ryu, C. & Heinzen, D. J. Molecules in a Bose-Einstein condensate. *Science* **287**, 1016–1019 (2000).
35. Heinzen, D. J., Wynar, R., Drummond, P. D. & Kheruntsyan, K. V. Superchemistry: dynamics of coupled atomic and molecular Bose-Einstein condensates. *Phys. Rev. Lett.* **84**, 5029–5033 (2000).
36. Mewes, M. O. *et al.* Bose-Einstein condensation in a tightly confining DC magnetic trap. *Phys. Rev. Lett.* **77**, 416–419 (1996).
37. Bradley, C. C., Sackett, C. A. & Hulet, R. G. Bose-Einstein condensation of lithium: observation of limited condensate number. *Phys. Rev. Lett.* **78**, 985–989 (1997).
38. Dalfovo, F., Giorgini, S., Pitaevskii, L. P. & Stringari, S. Theory of Bose-Einstein condensation in trapped gases. *Rev. Mod. Phys.* **71**, 463–512 (1999).
39. Pitaevskii, L. P. Vortex lines in an imperfect Bose gas. *Sov. Phys. JETP* **13**, 451–454 (1961).
40. Gross, E. P. Structure of a quantized vortex in boson systems. *Nuovo Cimento* **20**, 454–477 (1961).
41. Bogoliubov, N. N. On the theory of superfluidity. *J. Phys. (Moscow)* **11**, 23 (1947).
42. Gardiner, C. W. Particle-number-conserving Bogoliubov method which demonstrates the validity of the time-dependent Gross-Pitaevskii equation for a highly condensed Bose gas. *Phys. Rev. A* **56**, 1414–1423 (1997).
43. Castin, Y. & Dum, R. Low-temperature Bose-Einstein condensates in time-dependent traps: beyond the U(1) symmetry-breaking approach. *Phys. Rev. A* **57**, 3008–3021 (1998).
44. Burnett, K. in *Bose-Einstein Condensation in Atomic Gases* (eds Inguscio, M., Stringari, S. & Wieman, C. E.) 265–285 (IOS Press, Amsterdam, 1999).
45. Zaremba, E., Nikuni, T. & Griffin, A. Dynamics of trapped Bose gases at finite temperature. *J. Low Temp. Phys.* **116**, 277–345 (1999).
46. Miesner, H.-J. *et al.* Bosonic stimulation in the formation of a Bose-Einstein condensate. *Science* **279**, 1005–1007 (1998).

47. Köhl, M., Hänsch, T. W. & Esslinger, T. Growth of Bose-Einstein condensates from thermal vapor. Preprint cond-mat/0106642 at <http://xxx.lanl.gov> (2001).
48. Moss, S. *Formation and Decay of a Bose-Einstein Condensate in Atomic Hydrogen*. Thesis, MIT (2001).
49. Gardiner, C. W., Lee, M. D., Ballagh, R. J., Davis, M. J. & Zoller, P. Quantum kinetic theory of condensate growth: comparison of experiment and theory. *Phys. Rev. Lett.* **81**, 5266–5269 (1998).
50. Kocharovskiy, V. V., Scully, M. O., Zhu, S.-Y. & Zubairy, M. S. Condensation of N bosons. II. Nonequilibrium analysis of an ideal Bose gas and the laser phase-transition analogy. *Phys. Rev. A* **61**, 023609-1–023609-20 (2000).
51. Walser, R., Williams, J., Cooper, J. & Holland, M. Quantum kinetic theory for a condensed bosonic gas. *Phys. Rev. A* **59**, 3878–3889 (1999).
52. Kagan, Y. & Svistunov, B. V. Evolution of correlation properties and appearance of broken symmetry in the process of Bose-Einstein condensation. *Phys. Rev. Lett.* **79**, 3331–3334 (1997).
53. Bijlsma, M. J., Zaremba, E. & Stoof, H. T. C. Condensate growth in trapped Bose gases. *Phys. Rev. Lett.* **62**, 063609-1–063609-16 (2000).
54. Sackett, C. A., Gerton, J. M., Welling, M. & Hulet, R. G. Measurement of collective collapse in a Bose-Einstein condensate with attractive interactions. *Phys. Rev. Lett.* **82**, 876–879 (1999).
55. Donley, E. A. *et al.* Dynamics of collapsing and exploding Bose-Einstein condensates. *Nature* **412**, 295–299 (2001).
56. Hodby, E., Maragò, O. M., Hechenblaikner, G. & Foot, C. J. Experimental observation of Beliaev coupling in a Bose-Einstein condensate. *Phys. Rev. Lett.* **86**, 2196–2199 (2001).
57. Burger, S., Bongs, K., Dettmer, S., Ertmer, W. & Sengstock, K. Dark solitons in Bose-Einstein condensates. *Phys. Rev. Lett.* **83**, 5198–5201 (1999).
58. Denschlag, J. *et al.* Generating solitons by phase engineering of a Bose-Einstein condensate. *Science* **287**, 97–101 (2000).
59. Wyatt, A. F. G. Evidence for a Bose-Einstein condensate in liquid ⁴He from quantum evaporation. *Nature* **391**, 56–59 (1998).
60. Raman, C. *et al.* Evidence for a critical velocity in a Bose-Einstein condensed gas. *Phys. Rev. Lett.* **83**, 2502–2505 (1999).
61. Onofrio, R. *et al.* Observation of superfluid flow in a Bose-Einstein condensed gas. *Phys. Rev. Lett.* **85**, 2228–2231 (2000).
62. Burger, S., Cataliotti, F. S., Fort, C., Minardi, F. & Inguscio, M. Superfluid and dissipative dynamics of a Bose-Einstein condensate in a periodic optical potential. *Phys. Rev. Lett.* **86**, 4447–4450 (2001).
63. Cataliotti, F. S. *et al.* Josephson junction arrays with Bose-Einstein condensates. *Science* **293**, 843–846 (2001).
64. Guéry-Odelin, D. & Stringari, S. Scissors mode and superfluidity of a trapped Bose-Einstein condensed gas. *Phys. Rev. Lett.* **83**, 4452–4455 (1999).
65. Maragò, O. M. *et al.* Observation of the scissors mode and evidence for superfluidity of a trapped Bose-Einstein condensed gas. *Phys. Rev. Lett.* **84**, 2056–2059 (2000).
66. Williams, J. E. & Holland, M. J. Preparing topological states of a Bose-Einstein condensate. *Nature* **401**, 568–572 (1999).
67. Matthews, M. R. *et al.* Vortices in a Bose-Einstein condensate. *Phys. Rev. Lett.* **83**, 2498–2501 (1999).
68. Madison, K. W., Chevy, F., Wohlleben, W. & Dalibard, J. Vortex formation in a stirred Bose-Einstein condensate. *Phys. Rev. Lett.* **84**, 806–809 (2000).
69. Abo-Shaeer, J. R., Raman, C., Vogels, J. M. & Ketterle, W. Observation of vortex lattices in Bose-Einstein condensates. *Science* **292**, 476–479 (2001).
70. Haljan, P. C., Coddington, I., Engels, P. & Cornell, E. A. Driving Bose-Einstein-condensate vorticity with a rotating normal cloud. *Phys. Rev. Lett.* **87**, 210403-1–210403-4 (2001).
71. Anderson, B. P. *et al.* Watching dark solitons decay into vortex rings in a Bose-Einstein condensate. *Phys. Rev. Lett.* **86**, 2926–2929 (2001).
72. Hodby, E., Hechenblaikner, G., Hopkins, S. A., Maragò, O. M. & Foot, C. J. Vortex nucleation in Bose-Einstein condensates in an oblate, purely magnetic potential. *Phys. Rev. Lett.* **88**, 010405-1–010405-4 (2002).
73. Fetter, A. L. & Svidzinsky, A. A. Vortices in a trapped dilute Bose-Einstein condensate. *J. Phys. Condens. Matter* **13**, R135–R194 (2001).
74. Feder, D. L., Svidzinsky, A. A., Fetter, A. L. & Clark, C. W. Anomalous modes drive vortex dynamics in confined Bose-Einstein condensates. *Phys. Rev. Lett.* **86**, 564–567 (2001).
75. Madison, K. W., Chevy, F., Bretin, V. & Dalibard, J. Stationary states of a rotating Bose-Einstein condensate: routes to vortex nucleation. *Phys. Rev. Lett.* **86**, 4443–4446 (2001).
76. Sinha, S. & Castin, Y. Dynamic instability of a rotating Bose-Einstein condensate. *Phys. Rev. Lett.* **87**, 190402-1–190402-4 (2001).
77. Dalfovo, F. & Stringari, S. Shape deformations and angular-momentum transfer in trapped Bose-Einstein condensates. *Phys. Rev. A* **63**, 011601-1–011601-4 (2001).
78. Anglin, J. R. Local vortex generation and the surface mode spectrum of large Bose-Einstein condensates. *Phys. Rev. Lett.* **87**, 240401-1–240401-4 (2001).
79. Anglin, J. R. Vortices near surfaces of Bose-Einstein condensates. Preprint cond-mat/0110389 at <http://xxx.lanl.gov> (2001).
80. Winiecki, T., Jackson, B., McCann, J. F. & Adams, C. S. Vortex shedding and drag in dilute Bose-Einstein condensates. *J. Phys. B* **33**, 4069–4078 (2000).
81. Feder, D. L. & Clark, C. W. Superfluid-to-solid crossover in a rotating Bose-Einstein condensate. *Phys. Rev. Lett.* **87**, 190401-1–190401-4 (2001).
82. Hall, D. S., Matthews, M. R., Ensher, J. R., Wieman, C. E. & Cornell, E. A. Dynamics of component separation in a binary mixture of Bose-Einstein condensates. *Phys. Rev. Lett.* **81**, 1539–1542 (1998).
83. Stenger, J. *et al.* Spin domains in ground-state Bose-Einstein condensates. *Nature* **396**, 345–348 (1998).
84. Busch, T. & Anglin, J. R. Wave-function monopoles in Bose-Einstein condensates. *Phys. Rev. A* **60**, R2669–R2672 (1999).
85. Stoof, H. T. C., Vliegen, E. & Al Khawaja, U. Monopoles in an antiferromagnetic Bose-Einstein condensate. *Phys. Rev. Lett.* **87**, 120407-1–120407-4 (2001).
86. Ho, T.-L. Spinor Bose condensates in optical traps. *Phys. Rev. Lett.* **81**, 742–745 (1998).
87. Al Khawaja, U. & Stoof, H. Skyrmions in a ferromagnetic Bose-Einstein condensate. *Nature* **411**, 918–920 (2001).
88. Ruostekoski, J. & Anglin, J. R. Creating vortex rings and three-dimensional skyrmions in Bose-Einstein condensates. *Phys. Rev. Lett.* **86**, 3934–3937 (2001).
89. Law, C. K., Pu, H. & Bigelow, N. P. Quantum spins mixing in spinor Bose-Einstein condensates. *Phys. Rev. Lett.* **81**, 5257–5261 (1998).
90. Görlitz, A. *et al.* Realization of Bose-Einstein condensates in lower dimensions. *Phys. Rev. Lett.* **87**, 130402-4–130402-4 (2001).
91. Ketterle, W. & van Druten, N. J. Bose-Einstein condensation of a finite number of particles trapped in one or three dimensions. *Phys. Rev. A* **54**, 656–660 (1996).
92. Petrov, D. S., Shlyapnikov, G. V. & Walraven, J. T. M. Regimes of quantum degeneracy in trapped 1D gases. *Phys. Rev. Lett.* **85**, 3745–3749 (2000).
93. Dettmer, S. *et al.* Observation of phase fluctuations in elongated Bose-Einstein condensates. *Phys. Rev. Lett.* **87**, 160406-1–160406-4 (2001).
94. Monien, H., Linn, M. & Elstner, N. Trapped one-dimensional Bose gas as a Luttinger liquid. *Phys. Rev. A* **58**, R3395–R3398 (1998).
95. Olshanii, M. Atomic scattering in presence of an external confinement and a gas of impenetrable bosons. *Phys. Rev. Lett.* **81**, 938–941 (1998).
96. Bouyer, P. & Kasevich, M. A. Heisenberg-limited spectroscopy with degenerate Bose-Einstein gases. *Phys. Rev. A* **56**, R1083–R1086 (1997).
97. Orzel, C., Tuchman, A. K., Fenselau, M. L., Yasuda, M. & Kasevich, M. A. Squeezed states in a Bose-Einstein condensate. *Science* **291**, 2386–2389 (2001).
98. Jaksch, D., Bruder, C., Cirac, J. I., Gardiner, C. W. & Zoller, P. Cold bosonic atoms in optical lattices. *Phys. Rev. Lett.* **81**, 3108–3111 (1998).
99. Greiner, M., Mandel, O., Esslinger, T., Hänsch, T. W. & Bloch, I. Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms. *Nature* **415**, 39–44 (2002).
100. Heinzen, D. J. in *Bose-Einstein Condensation in Atomic Gases* (eds Inguscio, M., Stringari, S. & Wieman, C. E.) 351–390 (IOS Press, Amsterdam, 1999).

Acknowledgements

We are indebted to the whole BEC group at MIT for discussions. Our work is supported by NSF, ONR, ARO, NASA, and the David and Lucile Packard Foundation.

Nonlinear and quantum atom optics

S. L. Rolston & W. D. Phillips

Atomic Physics Division, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, Maryland 20899-8424, USA
(e-mail: srolston@nist.gov)

Coherent matter waves in the form of Bose–Einstein condensates have led to the development of nonlinear and quantum atom optics — the de Broglie wave analogues of nonlinear and quantum optics with light. In nonlinear atom optics, four-wave mixing of matter waves and mixing of combinations of light and matter waves have been observed; such progress culminated in the demonstration of phase-coherent matter-wave amplification. Solitons represent another active area in nonlinear atom optics: these non-dispersing propagating modes of the equation that governs Bose–Einstein condensates have been created experimentally, and observed subsequently to break up into vortices. Quantum atom optics is concerned with the statistical properties and correlations of matter-wave fields. A first step in this area is the measurement of reduced number fluctuations in a Bose–Einstein condensate partitioned into a series of optical potential wells.

The advent of the laser in 1960 began a new era in optics, eventually leading to numerous technological innovations, from laser surgery to CD-ROMs. Laser light has a combination of high coherence and high intensity that had been previously unattainable. These properties represent a significant difference from earlier light sources, and new kinds of phenomena became possible. Among them were nonlinear optical phenomena and the production of non-classical (that is, quantum) light. The production of atomic-gas Bose–Einstein condensates (BECs)^{1,2} brought a similar change in the optics of matter waves (atom optics).

One of the first, qualitatively new experiments to follow the appearance of the laser was second harmonic generation, or frequency doubling³. An intense pulse of red laser light irradiated a transparent crystal and the emerging pulse included a small amount of blue light, with twice the frequency (half the wavelength) of the red light. The blue light arose because the crystal responded nonlinearly to the electric field of the incident laser (the index of refraction depends on the light intensity). This and other nonlinear phenomena have made nonlinear optics an important and exciting field of research for the past 40 years⁴, with applications in physics, chemistry and biology.

The inverse process to second harmonic generation (or sum frequency generation in the case of two input frequencies) is parametric down-conversion⁵. Here, photons from laser light incident on a crystal are converted into twin photons whose frequency and momentum sum to that of the incident photons. The twins are strongly correlated in that one always accompanies the other, born simultaneously (within the inverse of their frequency bandwidth), with complementary frequency and direction. The number of photons emitted in complementary directions is identical, representing light beams more alike than any ordinary beams could ever be. In contrast, a laser beam incident on a perfect 50/50 beam splitter produces streams of photons whose numbers are similar only to within the statistical uncertainty of Poisson counting statistics (the beam splitter ‘flips a coin’ for each photon that passes through).

This equality of photon numbers produced in parametric down-conversion is impossible with any ‘classical’ light field (that is, one that could be produced by a classical distribution of oscillating charges and currents). The ‘squeezing’ of relative fluctuations between the light fields, and other examples of non-classical light, are at the heart of

much of modern quantum optics. This field has done much to further our understanding of quantum mechanics and the limitations it places on measurement.

Only a few years ago, the idea that atoms might exhibit such phenomena as nonlinear mixing to produce new atom waves, or squeezing of fluctuations below the usual quantum limits, was somewhat fanciful. Today, BECs and their nonlinear atom–atom interactions have made these ideas a reality for atom optics, just as the laser did for photon optics.

Coherent atom optics

Light from a laser is ‘coherent’, unlike the light from, for example, an incandescent lamp. This coherence is both temporal (the light is monochromatic, or single frequency) and spatial (the light appears to come from a single, point-like source rather than from an extended source). A BEC is similarly coherent. Bose–Einstein condensation occurs when a gas of bosonic particles (atoms in our case) is made cold enough and dense enough that the de Broglie wavelength of the atoms is comparable to the average interparticle spacing. The resulting phase transition puts a macroscopic fraction of the total number of atoms in the lowest available state of centre-of-mass motion. (This is strictly true in the limit of vanishing atom–atom interactions, but remains a useful picture for the weakly interacting BECs described here.) Just as a large number of photons in a single mode of the electromagnetic field characterizes a laser, a large number of atoms in a single quantum state characterizes a BEC. In the limit of no interactions, a BEC can be thought of as many atoms each with the same single-particle wavefunction. With weak interactions each particle is, to a good approximation, in an identical effective Hartree wavefunction, which accounts for the atom–atom interactions with a mean-field approximation. The quantum phase is uniform across the condensate. This single state and uniform phase are equivalent to the coherence of laser light.

Another feature that distinguishes a laser from a thermal (light bulb) source is the statistics of the fluctuations of the intensity of the light emitted. A laser emits photons in what is known as a coherent state, where the detection of one photon tells us nothing about the probability of detecting the next photon. Thermal light, on the other hand, exhibits photon ‘bunching’ — if a photon is detected there is a higher probability of detecting a second one soon after.

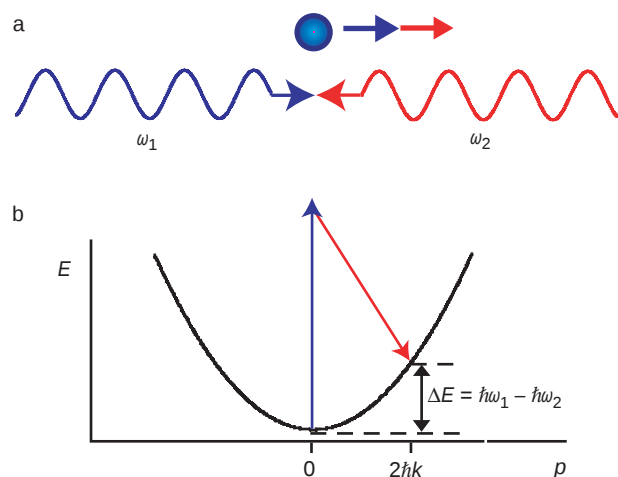


Figure 1 The fundamental process of momentum transfer between laser fields and atoms involved in Bragg scattering. **a**, An atom absorbs a photon from one laser beam (ω_1), and is stimulated to emit it into the other (ω_2). **b**, The dispersion relation for a free atom is shown as a parabola, $E = p^2/2m$. The difference in momentum of the two photons is the change in the momentum of the atom, and the difference in the photon energy is the change in kinetic energy of the atom.

Light from a light bulb is noisier than laser light. A BEC has the same property as a laser, which has been verified with a collision experiment⁶ that probed the probability of three atoms being near each other (the spatial analogue of the temporal measure of statistics in a laser).

The Schrödinger equation, which governs the way that atoms behave, is very similar in structure to the classical wave equation that governs light waves. Therefore, it is no surprise that it is often possible to treat matter waves in much the same way as light waves. Atom optics is the manipulation of atoms waves in a manner analogous to the manipulation of light (with lenses, beam splitters, diffraction gratings, and so on). Coherent atom optics should preserve the coherence of the atoms, and momentum transfer between coherent light and coherent atoms provides the coherent manipulation needed. To assure only coherent interactions, experimenters detune the laser frequencies far enough from atomic resonance that photon absorption followed by spontaneous emission (which would lead to decoherence) is negligible. Reflection, beam splitting, and diffraction of atoms can be thought of as changing the momentum of the atoms, from an initial value to a single final value, or to a superposition of momenta. One routine way in which this is done is known as Bragg diffraction⁷.

Bragg diffraction usually refers to X-ray scattering from crystals. Momentum and energy conservation can be satisfied only for the incidence angle where the scattering from each atomic plane constructively interferes, leading to specular reflection. In atom optics, the role of matter and light are reversed — the atomic wave scatters off a periodic standing light wave (optical lattice). In the case of a BEC, the standing wave is moving, such that as viewed in the moving frame of the lattice, the matter wave specularly reflects, reversing its velocity. Atomic Bragg scattering was first observed with a thermal atomic beam incident on a standing light wave⁸.

As illustrated in Fig. 1, Bragg-diffracted atoms coherently redistribute photons between two intersecting laser beams of different frequency, a process that produces a coherent change in the momentum and energy of the atoms. The atoms can be controllably left in a coherent superposition of the initial and final momentum states by choosing the appropriate laser intensity and pulse duration. Thus Bragg diffraction can operate as a mirror (when all atoms are

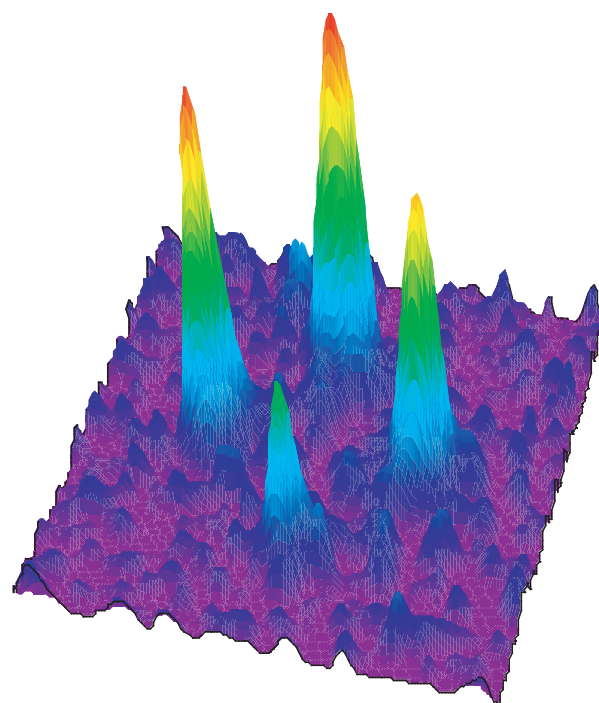


Figure 2 Results of four-wave mixing of atomic de Broglie waves¹⁸. Each peak represents atoms with different momenta observed after a period of free flight, which allows the different components to separate. The three largest peaks are the three initial momenta created from a Bose-Einstein condensate (BEC) by Bragg diffraction of atoms by light. The fourth wave, created by the nonlinear interaction between atoms, is the smallest of the peaks. The largest of the peaks has been amplified in the same mixing process that produces the fourth wave.

transferred to the new momentum state) or a beam splitter (when the atoms are in a superposition of two momentum states).

Bragg diffraction is accomplished using laser pulses long enough that the Fourier spread of photon energies is less than the change in atomic kinetic energy accompanying the momentum changes. For shorter pulses, coherent diffraction into multiple orders is possible⁹. Interferometric^{10–12} and other experiments^{13,14} have all confirmed the coherence of diffractive manipulation of BECs, and it is now a well-established tool to create specific matter wave states with well-defined momenta.

Four-wave mixing of atoms

Four-wave mixing of light¹⁵ is a higher-order extension of the second-order nonlinear process mentioned earlier that generates blue light from red light. In that process, two light fields E_{ω_1} and E_{ω_2} ($\omega_1 = \omega_2$ for frequency doubling) produce a polarization $P_{\omega_1 \pm \omega_2} = \chi_2 E_{\omega_1} E_{\omega_2}$ where χ_2 is the second-order susceptibility of the medium, representing its nonlinear nature. The polarization $P_{\omega_1 + \omega_2}$ radiates the sum frequency. In four-wave mixing, three fields produce a polarization $P_{\omega_1 \pm \omega_2 \pm \omega_3} = \chi_3 E_{\omega_1} E_{\omega_2} E_{\omega_3}$ with χ_3 being the third-order susceptibility. In a typical case, the polarization would radiate a fourth wave at $\omega_4 = \omega_1 + \omega_2 - \omega_3$, with wave vector $\mathbf{k}_4 = \mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_3$. These conditions reflect the requirements of energy and momentum conservation. A photon of frequency ω_4 results from the disappearance of photons at ω_1 and ω_2 and the stimulated emission of one at ω_3 .

The same four-wave mixing process can occur in a coherent sample of atoms, with three different de Broglie waves mixing to produce a fourth wave^{16,17}. In photon optics, the nonlinear susceptibility is a property of the medium. In atom optics, the nonlinearity comes for

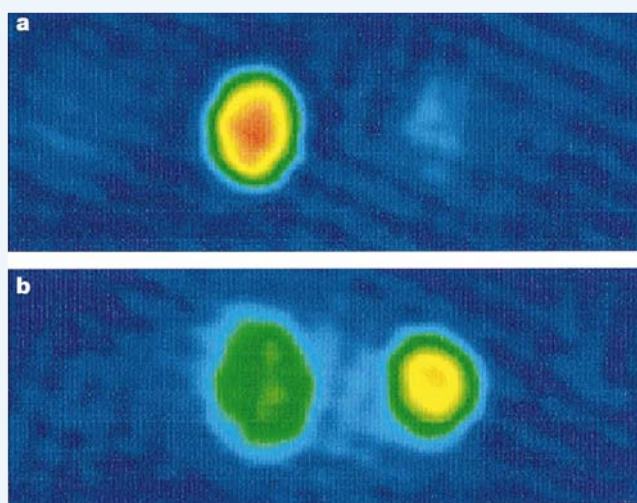


Figure 3 Matter-wave amplification¹¹. **a**, The momentum distribution of a rubidium condensate that was illuminated by Bragg pulses, producing a small (6%) amount of diffracted atoms at $2\hbar k$ (faintly visible to the right of the BEC). **b**, A BEC prepared as in **a** has been illuminated by a weak-probe laser beam. The small $2\hbar k$ component of the BEC seeds the amplifier, such that this momentum component is preferentially amplified. As much as 66% of the atomic population ends up with a momentum of $2\hbar k$.

free, from atom–atom interactions (coherent, elastic s-wave collisions; see review in this issue by Julienne and colleagues, pages 225–232). Although it may seem strange to speak of atom waves mixing to create a new atom wave, the Gross–Pitaevskii or nonlinear Schrödinger equation¹⁸ that describes the behaviour of a BEC is in fact quite similar to the equation that describes light in a nonlinear medium with a third-order susceptibility.

A different picture of four-wave mixing of light is that interfering light waves in a nonlinear medium produce a periodic variation of refractive index from which another wave Bragg diffracts. Atomic four-wave mixing can also be viewed this way: the interference of two of the atom waves form a grating, a standing wave or periodic variation of atomic density in the form of a stack of alternating planes of high and low density. The third atom wave can then Bragg-diffract from this set of planes, and the diffracted wave is the fourth, newly generated wave.

The experiment demonstrating four-wave mixing of atoms¹⁹ started with a BEC and successively applied pulses of intersecting laser beams to create two additional momentum states by Bragg diffraction. This preparation phase is so short that the resulting three momentum states (one of which is the initial $p=0$ state) are completely overlapping in space. The three waves interact through their atom–atom collisions, producing a fourth wave by the Bragg scattering of atoms by atoms. After a period of free flight, the momentum components separate and we see, in Fig. 2, the original three components and the new, fourth one.

As in optical four-wave mixing, the initial three momenta must be chosen carefully so the condition for energy and momentum conservation is satisfied ('phase matching'). When three initial momenta that did not satisfy the phase-matching condition were created, only the original three momentum components and not a fourth one were observed. In the experiment the amplitudes of the three initial momentum components were prepared to be equal, yet one of the peaks is clearly larger than the others in Fig. 2. This is because the mixing process creates a new wave, amplifies one of the original three waves, and de-amplifies the other two.

We expect the process of amplification to be correlated completely with the creation of the fourth wave (although this has yet to be confirmed experimentally). That is, for every atom put into the newly

generated fourth wave, exactly one atom should be added to the amplified wave. This correlation is reminiscent of the correlations present in parametric down-conversion in photon optics. It would be difficult to observe such correlations in the four-wave mixing experiment because of the large, uncorrelated background from the input-state wave being amplified. Nevertheless, it shows how non-classical correlations might arise from nonlinear interactions in atom optics, and is an example of amplification of matter waves.

An atom-optical analogue²⁰ to frequency doubling does exist — photoassociation of cold atoms into dimer molecules. Although photoassociation in a BEC has been observed, the coherent production of cold molecules remains an open question. This issue is discussed in the accompanying review article on ultracold collisions by Julienne and colleagues, pages 225–232.

Amplifiers

Four-wave mixing does not need to be confined solely to atom waves, or solely to light fields. In fact, Bragg diffraction of atoms by light can be viewed as four-wave mixing, with two light waves and two matter waves. The three input waves are the two light beams and the initial condensate, and the fourth wave is the Bragg-diffracted matter wave. Just as in all-atomic four-wave mixing, two of the waves (in this case one laser field and one atom field) are de-amplified, one field (laser) is amplified, and a fourth wave (matter) is created. This amplification and de-amplification of the laser beams (photon redistribution) has not been observed directly for atomic Bragg diffraction, although photon redistribution in other contexts has been observed²¹.

The illumination of an elongated BEC with a single laser beam by researchers at the Massachusetts Institute of Technology (MIT) produced a surprise: a coherent scattering of the condensate into a new momentum state and the coherent, directed emission of scattered light²². This is again interpretable as four-wave mixing, where the inputs consisted of a light wave and two matter waves (the BEC and a recoiling atom wave created by the spontaneous emission of a single photon). When a photon was scattered along the long direction of the cigar-shaped condensate, the recoiling atom produced a Bragg grating that enhanced the probability of more emissions being scattered in this direction — a gain process for this mode, emphasizing the amplification present in four-wave mixing. This can be exploited to build a matter-wave amplifier (four-wave mixing with inputs of two atom waves and one laser wave), which will amplify an input de Broglie wave, rather than one that was generated spontaneously.

Soon after these experiments, both the MIT group and a group at the University of Tokyo demonstrated phase-coherent matter-wave amplifiers^{11,12}. Figure 3 shows the results of the Tokyo experiment, where a weak Bragg pulse created a small-amplitude de Broglie wave with a well-defined momentum. This wave interfered with the remaining condensate, creating a matter-wave grating. This grating then Bragg diffracted an applied laser pulse, and the recoiling atoms were created with the same momentum as the small initial wave (simply a consequence of energy–momentum conservation). The recoiling wave constructively interfered with the initial wave, creating a deeper matter grating, which further increased the scattering into this mode. The result was an amplification of the input wave, by as much as a factor of 30 in the MIT experiment. The process should be fully coherent, and both groups were able to verify this by creating atom interferometers. The observed interference between the amplified wave and reference matter waves created with additional Bragg pulses confirmed that these were indeed phase-coherent matter-wave amplifiers. Because amplification is at the heart of an optical laser, we can expect that matter-wave amplifiers may find application in increasing the brightness of atom lasers (see Box 1).

We have presented a physical interpretation of four-wave mixing as arising from scattering from gratings, whether made of atoms or light. A completely equivalent interpretation, but one seemingly very different, is to view four-wave mixing and matter-wave amplification as a consequence of Bose stimulation — the quantum statistical

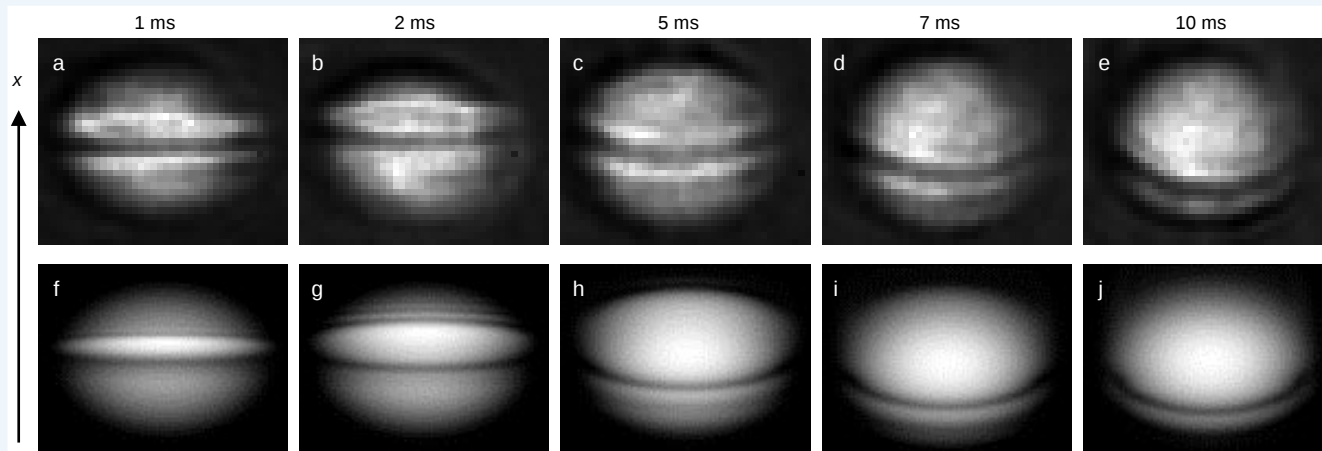


Figure 4 Dark solitons produced in a BEC with phase imprinting²⁸. **a–e**, Experimental images of the integrated atomic density; **f–j**, solutions of the three-dimensional Gross–Pitaevskii equation, after a $\sim 1.5\pi$ phase step was imprinted on the upper half

of the condensate. A positive density disturbance (sound waves) moved rapidly upward, and a dark soliton moved oppositely at much less than the speed of sound. The curvature is a result of the inhomogeneous atomic density.

property that states that the scattering of bosons into a particular mode is proportional to the number of bosons already in that mode. This seemingly special statistical property of bosons is in a sense just a classical wave property — a valid physical picture of the stimulated emission of light is simply the interference of a spontaneously generated wave with the incident wave. This constructive interference is called ‘Bose stimulation’ when we refer to matter waves, and stimulated emission when we refer to light waves. But the issue is even subtler. Two recent papers^{23,24} have shown that it is possible to do four-wave mixing (or something that looks almost indistinguishable) with fermions, where there is explicitly no stimulation factor, because only one fermion can occupy one mode. The answer to this paradox seems to lie in constructive multi-particle interferences, where a piece of the total Fermi wavefunction can have the same symmetry properties as a Bose wavefunction, and produce roughly the same enhancement of scattering into a well-defined direction.

Solitons

The intensity-dependent index of refraction of a nonlinear medium not only produces phenomena such as wave mixing and harmonic generation, but also can affect the propagation of a pulse of light. If light passes through a region with a spatially varying index of refraction, it can get deflected, distorted or even focused or defocused. If the light itself causes the variation in the index, this is a nonlinear process.

One such phenomenon that is being implemented commercially in optical communications is known as a soliton, a wave packet that can propagate for long distances in a nonlinear medium without spreading. Solitons (solitary waves) were first discovered²⁵ in 1834 by John Scott Russell, a Scottish engineer who followed on horseback a single ‘hump’ of water that travelled for more than a mile in a canal without spreading. Solitons are ubiquitous in many nonlinear systems, where they represent stable solutions of the nonlinear wave equation that governs propagation (we will not distinguish here between solitons and solitary waves). The type of soliton observed by Russell and used in fibre communications is a ‘bright’ soliton. It consists of a pulse of light with a specific envelope that does not change in time as it moves, because the intensity-dependent index of refraction exactly compensates dispersion (different frequencies travelling at different speeds).

The nonlinear equation that describes atoms in a BEC has the same structure as the nonlinear wave equation that governs the envelope of a pulse of light propagating in a fibre. But the sign of the nonlinear term is opposite (for atoms with a repulsive interaction),

so that the stable solutions are dark solitons (also observed in fibres for certain parameters, but less useful for communications). A dark soliton consists of a notch in the density profile, with a particular shape that again allows for propagation without spreading. It also has a characteristic step in the phase across the notch. Solitons in a BEC have another distinguishing characteristic — they propagate at a velocity of less than the speed of sound. The speed of propagation is inversely proportional to the depth of the notch, such that if the notch extends all the way to zero density, the soliton is ‘black’ and is stationary.

Two groups, at the National Institute of Standards and Technology (NIST) in Gaithersburg, and the University of Hanover in Germany, set out to create and observe solitons in BECs^{26,27}. Both groups generated dark solitons by exploiting the characteristic phase step. They applied a short pulse of laser light that illuminated one half of the BEC. This light, tuned far from atomic resonance, produced a phase shift on the half of the BEC that was illuminated, creating a phase step across the BEC. After the laser pulse the BEC evolved for varying lengths of time, and the density profile was measured with absorption imaging.

The NIST results presented in Fig. 4 show that a dark notch formed and then propagated downwards. The notch could be definitively identified as a soliton because it travelled opposite the direction of momentum transfer from the gradient force of the illuminating laser beam (in the narrow region where the laser beam went from bright to dark, an optical force was exerted on the atoms), and its propagation velocity was less than the speed of sound. They observed that the deeper the notch, the slower the soliton propagated, as expected. The observed curvature of the solitons was a consequence of the density profile of the BEC, which meant that the speed of sound and soliton speed would also have a spatial variation. Also shown in Fig. 4 are solutions of the three-dimensional Gross–Pitaevskii equation, showing excellent agreement with the experimental results.

One of the interesting questions is what happens when the soliton propagates to the edge of the condensate. As the density decreases, does the notch remain a constant fraction of the local density, or does its depth remain constant until the bottom of the notch hits zero density, and it becomes a stationary black soliton? The Gross–Pitaevskii equations showed that the latter was in fact what happened²⁸, and that this led to a break-up of the solitons into vortices and vortex rings.

Researchers at Harvard University observed this recently²⁹ in an experiment where solitons were created in a long condensate using a

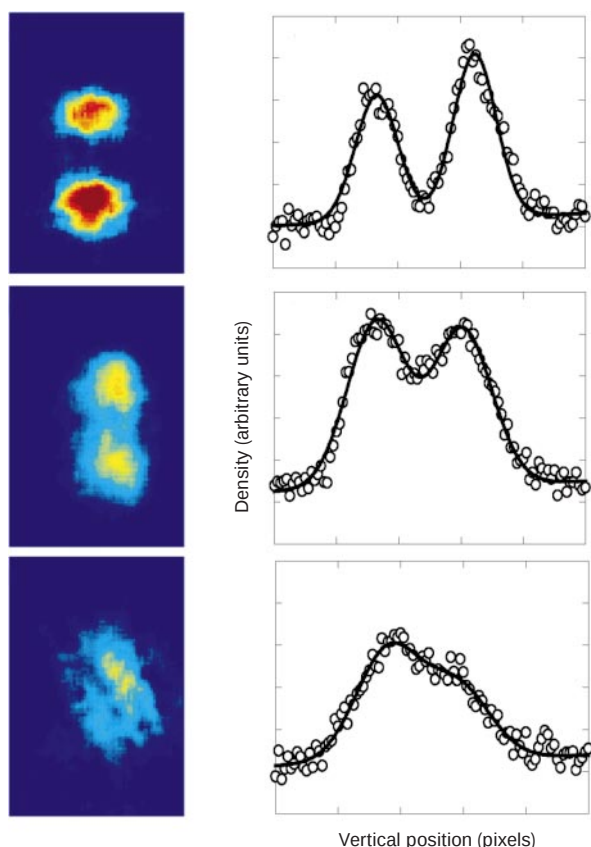


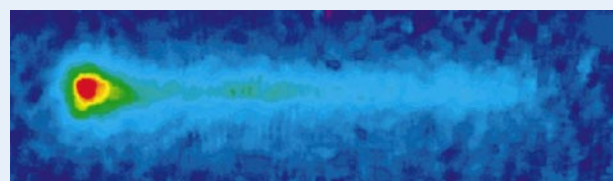
Figure 5 Loss of phase coherence as the atom-number distribution is squeezed in an optical lattice³⁴. The three curves correspond to optical lattices of depth 7.2, 18 and 44 E_R (E_R is the photon recoil energy). At the lowest depth, tunnelling is still prevalent, and there is phase coherence across the lattice sites, and a clearly visible diffraction pattern. For the deepest lattice, the number fluctuations in the wells are reduced, which scrambles the phase and washes out the interference pattern.

different method. They had shown previously³⁰ that by making an index of refraction that varied significantly over a very narrow velocity range (using laser light to couple three-level atomic systems), they could significantly alter the propagation of light through a BEC. It is possible to dramatically ‘slow’ light, and even to store it in the atomic media, in a way that makes it seem that the light has stopped. Using these techniques, they engineered a ‘light roadblock’ by illuminating only half of the BEC with the appropriate laser frequency that created the special coupling between the atomic levels. At this roadblock, there is a notch in the density of atoms in one state, and a peak in the density of the other state. This localized defect then creates up to five solitons (the number created depends on the strength of the defect). They further observed the break-up of the solitons through the so-called snake instability into vortex pairs. Similar examples of soliton break-up have been observed³¹ by researchers at JILA, an interdisciplinary research institute in Boulder, Colorado.

Solitons and vortices represent stable topological solutions of the nonlinear Schrödinger equation. (Vortices in BECs created directly through rotation have been studied extensively and are covered in the accompanying review article by Anglin and Ketterle, pages 211–218.) The rich dynamics evident in the evolution from solitons to vortices arises because solitons are the stable solutions for one dimension, whereas vortices are the stable three-dimensional features. Further studies with BECs can continue to probe these dynamics in an experimentally flexible system, where it will be possible to change the

Box 1 Atom lasers

Optical lasers produce photon beams by extracting coherent light stored in the laser cavity through a partially transmitting mirror. In the same way, an atom laser extracts coherent atoms from a Bose–Einstein condensate (BEC) as an output atomic beam. Here the BEC adopts the role of the intracavity field, the trap acts as the cavity, and various mechanisms have been used as output couplers. In the first atom laser⁴⁰, a radio-frequency pulse induced a fraction of the trapped atoms to make a transition to an untrapped state, whereupon those atoms fell out of the trap as a result of gravity. Other atom lasers have also used radio-frequency transitions⁴¹ or gravity-induced tunnelling⁴². In a sense, the first atomic-gas BEC¹ demonstrated a ‘cavity dumped’ atom laser, simply by turning off the trap and letting all the atoms escape. The figure below is the output beam of an atom laser⁴³ in which a two-photon optical Raman transition takes atoms from a trapped to an untrapped state, giving them a momentum kick from the two photons in the process. Although atom lasers show an analogy with photon lasers, they have yet to progress far beyond the demonstration stage.



Box 1 Figure An atom laser created with optical Raman pulses. The image of atomic density as taken after 140 1- μ s optical Raman pulses have been applied to a sodium BEC. The Raman pulses transfer the atoms to the $m=0$ Zeeman sublevel, where they no longer feel the magnetic trap, and also impart a velocity of 6 cm s⁻¹ (from two photon recoils). This produces a quasi-continuous atom laser beam, as the size of each output pulse is 30 μ m, while the separation between the peaks of sequential pulses is only 3 μ m. The horizontal dimension of the figure is $\sim$ 1.5 mm.

nature of the system (for example, change the dimensionality by changing the shape of the BEC) in a way not available in many other nonlinear systems.

Squeezing of atom numbers

Nonlinear optics has given us the ability to create non-classical states of light, such as squeezed states where the noise characteristics in one degree of freedom can be less than the so-called ‘standard quantum limit’ given by the Heisenberg uncertainty principle. Squeezing does not invalidate the uncertainty principle, but uses it by increasing the noise in the degree of freedom that is uninteresting, while reducing it in the relevant one. A recent experiment³² by researchers at Yale University has now shown that the nonlinearity in a BEC may also produce non-classical configurations of atoms. This experiment looked at the partitioning of a BEC into a set of atom clouds in about 12 wells of an optical lattice. Imagine that we took a BEC with N atoms and split it in half, putting half in one well, and the other half in another well. We furthermore assume (quite reasonably) that each atom has a 50% probability of landing in either the left or the right well. What should we expect for the distribution of atoms between the wells? Although the mean number in each well will be $N/2$, there will be an uncertainty of $\sqrt{N}/2$ (it is extremely unlikely to get exactly 500,000 heads if a coin is flipped 10⁶ times). This uncertainty in number is intrinsic in physics, and is a source of noise in many experimental situations, such as the fundamental noise (called shot-noise) on a laser beam. Classically, this noise is unavoidable.

The Yale group showed that with nonlinearities, this noise limit may be reduced significantly. They loaded a BEC of 10^5 rubidium atoms into an optical lattice, whose strength increased slowly in time (it took 200 ms to reach its final well depth). Early in the loading the atoms resided in a shallow optical lattice and, because they could easily tunnel from well to well, each well did not contain a well-defined number of atoms. But the atom–atom interaction responsible for the nonlinear term in the Gross–Pitaevskii equation also means there is a cost in energy when two atoms get close to one another. Having many atoms in one well is energetically unfavourable. Clearly the lowest energy state of this system would be to have N/M atoms in each of M wells. By raising the lattice intensity slowly enough, the system could start in the lowest possible energy state (the BEC) and remain in the lowest energy state (adiabatic following), which would evolve into an equal partitioning of the atoms into the wells of the lattice.

Because of the uncertainty principle, we know that whenever we reduce uncertainty in some variable (here the number of atoms in the wells), we must increase it in some other variable. It is not obvious what that other variable is here, but the answer can be found by appealing to solid-state physics. The lowest energy state in a periodic potential is the ground state band with a quasimomentum of zero, which corresponds to a state completely delocalized across the lattice with a uniform phase. If we consider an atom localized in a specific well, this can be written as a superposition of all quasimomenta, that is, all phases. So we see that forcing atoms to be in specific wells introduces an uncertainty into the phase of the wavefunction (this number–phase relationship is well known in quantum optics and optical squeezing).

The Yale group used this to infer the reduction of the fluctuations in atom number in the lattice wells. If an optical lattice potential is suddenly turned off, the result looks just like diffraction from multiple slits — each well corresponds to a slit and, if there is coherence across all the wells, a diffraction pattern is expected. What we see after a time-of-flight are peaks (in this case two) that correspond to momenta quantized in units of $2\hbar k$ (Fig. 5). The Yale experimenters observed the disappearance of the diffraction peaks as the number-squeezed state formed, because the phase in each well was becoming uncertain (scrambling the phase in front of each slit in a multiple-slit diffraction pattern would wash out the diffraction pattern). By measuring the change in the atomic diffraction pattern, and comparing this to theory, they inferred that the number fluctuations in each well were reduced from ~ 30 ($\sim \sqrt{N/M}$) to 2–3. A recent experiment³³ in a three-dimensional optical lattice has observed the complete disappearance of number fluctuations, as the system underwent a quantum phase transition into a so-called Mott insulator phase with exactly one atom per cell.

We can view this experiment as the first quantum atom-optics experiment, in analogy with quantum optics, a field concerned with the non-classical statistical properties of light. The experiment by the Yale University group shows that such effects exist with de Broglie waves. The recent observation of Bose–Einstein condensation in metastable helium^{34,35} opens the door to experiments that directly count atom number (it is easy to detect individual metastable atoms) allowing a direct probe of atom-number statistics. As mentioned above, four-wave mixing creates quantum correlations of atom number between the different waves. Such correlations are at the heart of much of our current ideas about quantum information, and ways to beat the standard quantum measurement limits. Considering the number of atoms (in a mode, or in a well, for example) as the projection of a fictitious collective spin, we can apply the concepts of spin squeezing³⁶ to these correlations. Recent work^{37–39} has shown that two-particle operators (acting on two atoms at a time, such as an elastic collision) when summed over all possible pairs of atoms, can produce spin squeezed and massively entangled states. What remains to be seen is whether such states will be robust enough to survive, be

measured, and be used for improved measurements or for quantum processing. Using the extensive tool kit of atom optics will provide many opportunities to create and probe the unusual many-particle states that quantum mechanics allows. □

- Anderson, M. H. *et al.* Observation of Bose–Einstein condensation in a dilute atomic vapor. *Science* **269**, 198–201 (1995).
- Inguscio, M., Stringari, S. & Wieman, C. (eds) *Bose–Einstein Condensation in Atomic Gases* (Int. School Phys. “Enrico Fermi” Course 140) (IOS Press, Amsterdam, 1999).
- Franken, P. A., Hill, A. E., Peters, C. W. & Weinreich, G. Generation of optical harmonics. *Phys. Rev. Lett.* **7**, 118–119 (1961).
- Evans, M. & Kielich, S. *Modern Nonlinear Optics* Vols 1–3 (Wiley, New York, 1997).
- Migdall, A. Correlated-photon metrology without absolute standards. *Phys. Today* **52**, 41–46 (1999).
- Burt, E. A. *et al.* Coherence, correlations and collisions: what one learns about Bose–Einstein condensates from their decay. *Phys. Rev. Lett.* **79**, 337–340 (1997).
- Kozuma, M. *et al.* Coherent splitting of Bose–Einstein condensed atoms with optically induced Bragg diffraction. *Phys. Rev. Lett.* **82**, 871–875 (1999).
- Martin, P. J., Oldaker, B. G., Miklich, A. H. & Pritchard, D. E. Bragg scattering of atoms from a standing light wave. *Phys. Rev. Lett.* **60**, 515–518 (1988).
- Ovchinnikov, Y. B. *et al.* Diffraction of a released Bose–Einstein condensate by a pulsed standing light wave. *Phys. Rev. Lett.* **83**, 284–287 (1999).
- Simsarian, J. E. *et al.* Imaging the phase of an evolving Bose–Einstein condensate wave function. *Phys. Rev. Lett.* **85**, 2040–2043 (2000).
- Kozuma, M. *et al.* Phase-coherent amplification of matter waves. *Science* **286**, 2309–2312 (1999).
- Inouye, S. *et al.* Phase-coherent amplification of atomic matter waves. *Nature* **402**, 641–644 (1999).
- Hagley, E. W. *et al.* Measurement of the coherence of a Bose–Einstein condensate. *Phys. Rev. Lett.* **83**, 3112–3115 (1999).
- Stenger, J. *et al.* Bragg spectroscopy of a Bose–Einstein condensate. *Phys. Rev. Lett.* **82**, 4569–4572 (1999).
- Shen, Y. R. *The Principles of Nonlinear Optics* (Wiley-Interscience, New York, 1984).
- Lens, G., Meystre, P. & Wright, E. W. Nonlinear atom optics. *Phys. Rev. Lett.* **71**, 3271–3274 (1993).
- Trippenbach, M., Band, Y. B. & Julienne, P. S. Four wave mixing in the scattering of Bose–Einstein condensates. *Opt. Express* **3**, 530–537 (1998).
- Dalfovo, F., Giorgini, S., Pitaevskii, L. P. & Stringari, S. Theory of Bose–Einstein condensation in trapped gases. *Rev. Mod. Phys.* **71**, 463–512 (1999).
- Deng, L. *et al.* Four-wave mixing with matter waves. *Nature* **398**, 218–220 (1999).
- Javanianen, J. & Mackie, M. Coherent photoassociation of a Bose–Einstein condensate. *Phys. Rev. A* **59**, R3186–R3189 (1999).
- Raithel, G., Phillips, W. D. & Rolston, S. L. Collapse and revivals of wave packets in optical lattices. *Phys. Rev. Lett.* **81**, 3615–3618 (1998).
- Inouye, S. *et al.* Superradiant Rayleigh scattering from a Bose–Einstein condensate. *Science* **285**, 571–574 (1999).
- Moore, M. G. & Meystre, P. Atomic four-wave mixing: fermions versus bosons. *Phys. Rev. Lett.* **86**, 4199–4202 (2001).
- Ketterle, W. & Inouye, S. Dose matter wave amplification work for fermions? *Phys. Rev. Lett.* **86**, 4203–4206 (2001).
- Russell, J. S. Report on waves. Report of the 14th meeting of the British Association for the Advancement of Science 331–390 (1844).
- Denschlag, J. *et al.* Generating solitons by phase engineering of a Bose–Einstein condensate. *Science* **287**, 97–100 (2000).
- Burger, S. *et al.* Dark solitons in Bose–Einstein condensates. *Phys. Rev. Lett.* **83**, 5198–5201 (1999).
- Feder, D. L. *et al.* Dark-soliton states of Bose–Einstein condensates in anisotropic traps. *Phys. Rev. A* **62**, 053606–1–053606–11 (2000).
- Dutton, Z., Budde, M., Stowe, C. & Hau, L. V. Observation of quantum shock waves created with ultra-slow light pulses in a Bose–Einstein condensate. *Science* **293**, 663–668 (2001).
- Liu, C., Dutton, Z., Behroozi, C. H. & Hau, L. V. Observation of coherent optical information storage in an atomic medium using halted light pulses. *Nature* **409**, 490–493 (2001).
- Anderson, B. P. *et al.* Watching dark solitons decay into vortex rings in a Bose–Einstein condensate. *Phys. Rev. Lett.* **86**, 2926–2929 (2001).
- Orzel, C., Tuchman, A. K., Feneclau, M. L., Yasuda, M. & Kasevich, M. A. Squeezed states in a Bose–Einstein condensate. *Science* **291**, 2386–2389 (2001).
- Greiner, M., Mandel, O., Esslinger, T., Hänsch, T. W. & Bloch, I. Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms. *Nature* **415**, 39–44 (2002).
- Robert, A. *et al.* A Bose–Einstein condensate of metastable atoms. *Science* **292**, 461–464 (2001).
- Pereira Dos Santos, F. *et al.* Bose–Einstein condensation of metastable helium. *Phys. Rev. Lett.* **86**, 3459–3462 (2001).
- Kitagawa, M. & Ueda, M. Spin squeezed states. *Phys. Rev. A* **47**, 5138–5143 (1993).
- Sørensen, A. & Mølmer, K. Spin-spin interaction and spin squeezing in an optical lattice. *Phys. Rev. Lett.* **83**, 2274–2277 (1999).
- Sørensen, A., Duan, L. M., Cirac, J. I. & Zoller, P. Many-particle entanglements with Bose–Einstein condensates. *Nature* **409**, 63–66 (2001).
- Helmerson, K. & You, L. Creating massive entanglement of Bose condensed atoms. *Phys. Rev. Lett.* **87**, 170402–1–170402–4 (2001).
- Mewes, M.-O. *et al.* Output coupler for Bose–Einstein condensed atoms. *Phys. Rev. Lett.* **78**, 582–585 (1997).
- Bloch, I., Hänsch, T. & Esslinger, T. Atom laser with a cw output coupler. *Phys. Rev. Lett.* **82**, 3008–3011 (1999).
- Anderson, B. P. & Kasevich, M. A. Macroscopic quantum interference from atomic tunnel arrays. *Science* **282**, 1686–1689 (1998).
- Hagley, E. W. *et al.* A well-collimated quasi-continuous atom laser. *Science* **283**, 1706–1709 (1999).

Acknowledgements

Work in this area at NIST-Gaithersburg was partly supported by the US Office of Naval Research, NASA and ARDA-NSA.

Quantum encounters of the cold kind

Keith Burnett*, Paul S. Julienne†, Paul D. Lett†, Eite Tiesinga† & Carl J. Williams†

*University of Oxford, Department of Physics, Clarendon Laboratory, Parks Road, Oxford, OX1 3PU, UK

†Atomic Physics Division, 100 Bureau Drive, MS 8423, National Institute of Standards and Technology, Gaithersburg, Maryland 20899-8423, USA (e-mail: paul.julienne@nist.gov)

Since the introduction of laser-cooling techniques for neutral atoms in the early 1980s, the study of collisional interactions between atoms and molecules has been extended to the regime of ultracold temperatures. With nanokelvin temperatures now attainable, our ability to probe the interactions, both experimentally and theoretically, has also progressed. Understanding of the subtle and often highly quantum-mechanical effects that are manifest at such low energies has advanced to the point where new precision measurements are matched by highly accurate theoretical calculations. Low-energy phenomena such as Bose–Einstein condensation and the photoassociation of atoms into bound molecules are now accurately described with no free parameters.

The behaviour of atoms and their interactions at ultracold temperatures is a fascinating area of study. These interactions and their effects distinguish them from those encountered in collisions at room temperature. The realization that these interactions would be both subtle and interesting began in the 1970s with studies¹ of spin-polarized hydrogen and long-range molecules, and expanded as laser cooling^{2–4} reached temperatures in the millikelvin and then microkelvin ranges (Fig. 1). With the advent of evaporative cooling⁵ and the production of atomic Bose–Einstein condensates (BECs; see ref. 6 and the review in this issue by Anglin and Ketterle, pages 211–218), we now require a detailed understanding of atomic interactions at nanokelvin temperatures. These applications have driven a tremendous growth of interest in the field.

Historical background

It was realized early on that the quantal nature of ultracold atomic interactions would have a profound role and provide a challenge to theorists and experimentalists in the field of collision dynamics. At the low energies involved, the precise nature of the interatomic forces — at ludicrously large distances from the point of view of ‘normal molecular physics’ — would have to be determined, requiring new ways to examine them. It was also evident that nuclear spin dynamics, usually irrelevant to collision dynamics, would complicate the analysis enormously. In fact, the term complicated does not do service to the range of new physics that the hyperfine interactions bring up in this regime.

It was also thought, correctly as we now know, that it would be possible to influence ultracold collision dynamics using external fields. This had been a subject of intense, but rather disappointing, study for the case of room-temperature collisions in the years before the development of laser cooling and trapping. The extremely low temperatures afforded by laser-cooling techniques meant that the range of energies participating in a collisional interaction would be extraordinarily narrow. Resonance phenomena could then be used to influence the whole of the gas rather than a small fraction of it, as is the case for room-temperature atomic collisions. The compression of the occupied phase space changes the nature of collisions from essentially incoherent thermal processes to coherent ones affecting the quantum dynamics of the system as a whole. This coherent dynamics

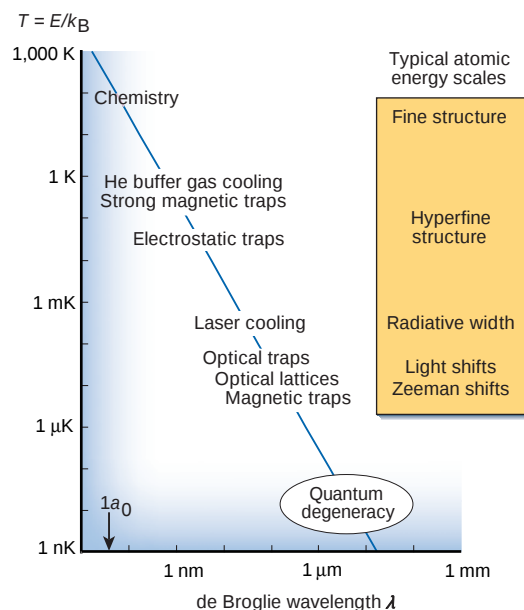
is subtle and complex and fully accessible to external probes and manipulation. In the 1980s it became apparent from the study of evaporative cooling of spin-polarized hydrogen that the precise nature of the interactions between alkali atoms would be crucial in determining whether or not it would be possible to achieve Bose–Einstein condensation with them.

Following on from the successes of laser- and evaporative-cooling techniques, several groups decided to study the nature of interactions between ultracold gases and develop the techniques needed to describe them¹. The development of new theoretical methods evolved along with new experimental techniques, and it is now an active area of study with strong connections between theoretical modelling and experimental work. Completely new spectroscopies have been developed, the precision of which reveals the limitations of existing models of molecular potentials. These spectroscopies have allowed refinement of these potentials, especially in the long-range region. The remarkable accuracy with which the specific atomic interactions can now be determined has made quantitative many-body theory for BECs possible. The fact that the interactions themselves depend sensitively on the states of the diatomic collision complex and can be tuned by the application of external fields means that we can produce ‘designer’ condensates. One of the most exciting applications of this tuning is switching the interactions from repulsive to attractive, and observing the collapse of a condensate⁷.

Laser cooling has been applied to alkali and alkaline earth species and excited metastable noble gas atoms. Quantum degeneracy by evaporative cooling in a dilute gas (Fig. 1; and see review in this issue by Anglin and Ketterle, pages 211–218) has been realized for the bosonic species ¹H, ⁷Li, ²³Na, ⁸⁷Rb, ⁸⁵Rb and metastable ³S₁ He, and for the fermionic species ⁶Li and ⁴⁰K. Although molecules cannot be laser cooled in a simple way, cold ground-state alkali dimer molecules of Li₂, Na₂, K₂, Rb₂ and Cs₂ have been produced by photoassociation of two already cold atoms^{8–14}. Although they are vibrationally excited, the molecules have a translational temperature commensurate with that of the source atoms. Helium buffer-gas cooling in a magnetic trap has also been applied to Cr and the CaH molecule¹⁵, and electrostatic beam slowing and trapping to the polar molecule ND₃ (ref. 16).

The possibility of doing quantum information processing with atoms in optical lattices has motivated the study of

Figure 1 Overview of neutral atomic and molecular cooling and trapping. The energy scale spans 12 orders of magnitude in kinetic energy E of atomic motion, expressed in temperature units $T = E/k_B$ (k_B is the Boltzmann constant). The length scale shows the corresponding de Broglie wavelength $\lambda = h/p$ (h is the Planck constant), where momentum $p = (2mE)^{1/2}$. The line to guide the eye is calculated using the mass of the ^{23}Na atom for m . Typical atomic dimensions are on the order of $1a_0 = 0.0529\text{nm}$, the Bohr radius of the H atom, whereas Bose–Einstein condensates can have dimensions on the order of $100\text{ }\mu\text{m}$. The figure also indicates typical orders of magnitude for familiar energy scales associated with atomic fine structure, hyperfine structure, natural radiative broadening and optical or magnetic energy shifts. Cooling requires a dissipative process to remove kinetic energy, whereas trapping requires a spatially dependent force to act on the atoms. Laser cooling^{2–4} typically reaches kinetic energies below 1 mK , where $k_B T$ is smaller than natural line widths, and atoms can be confined in a magneto-optical trap or optical trap. The latter uses a spatially dependent light shift attributable to a laser field to create a trapping potential for atoms. Similarly, a standing-wave light field can produce a periodic potential in space called an optical lattice⁵⁶ with individual lattice cells spaced by $\lambda_L/2$, where λ_L is the wavelength of the laser that makes the lattice. Atoms whose magnetic Zeeman shift decreases with decreasing field strength can be trapped in a region of space with a magnetic field minimum. Evaporative cooling⁵ of trapped atoms can dramatically lower the temperature and reach the limit of quantum degeneracy, where the phase space density ρ , defined as the number of particles per cubic thermal de Broglie wavelength, is of order unity. If the atoms are bosons, then Bose–Einstein condensation occurs when ρ reaches the critical value of 2.6 (ref. 6; and see review in this issue by Anglin and Ketterle, pages 211–218).



atoms interacting in optical lattices⁴ and on ‘atom chips’. Such systems might be used to entangle atomic states and to use them in quantum information processing and manipulation.

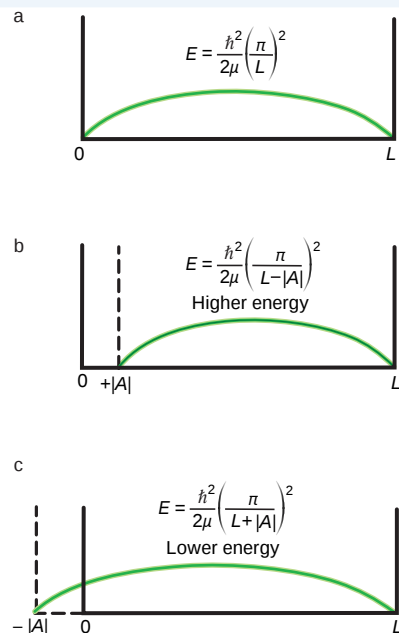
Ultracold interactions

The enormous, micrometre-scale de Broglie wavelengths ($1\text{ }\mu\text{m} \approx 1.9 \times 10^4 a_0$, where a_0 is the Bohr radius) of colliding atoms in these systems extend much further than atomic diameters (less than $10 a_0$). The scattering of a wave from an object that is much smaller than the wavelength produces predominantly isotropic (that is, s -wave) scattering. Box 1 explains the concept of the scattering length, which characterizes the scattering at very low collision energies attributable to complex, short-range atomic interactions. The

scattering length A can be much larger than the range of atomic interactions, but still smaller than the de Broglie wavelength. The precise value of the scattering length depends on the details of the interatomic potential, and very small variations in this potential can produce marked variations in the scattering length, even a change of sign.

Thus, each alkali isotope has its own unique ‘personality’ associated with the scattering lengths for each internal spin state. Because of this it is not possible to calculate sufficiently accurate scattering lengths using the conventional *ab initio* methods developed in atomic and molecular physics. Determining their values has required the development of new precision spectroscopies and sophisticated theoretical models. This has taken place along with the development

Figure 2 A heuristic argument for understanding why scattering length controls the energetics of particle interactions in a cold quantum gas. Consider a pair of atoms with reduced mass $\mu = m/2$, whose relative motion is confined to a box of length L . This length could represent the size of the trapping potential, for example. The box shows how the scattering length A , which characterizes the net effect of complicated short-range interactions, can be viewed as changing the boundary condition near $R = 0$ on the scattering wavefunction, which seems to have a node at $R = A$ instead of $R = 0$. This change in boundary condition leads to a change in kinetic energy due to particle interactions that is proportional to A . **a**, Wavefunction and quantized ground-state energy of the non-interacting pair in the box. **b**, Change in wavefunction and energy if the left-hand edge of the box is moved to the right by a distance $|A|$. **c**, Corresponding change if the edge of the box is moved to the left by distance $|A|$. This movement of the left-hand edge of the box resets the boundary condition and leads to a change in energy relative to the case in panel **a**. Simple algebra shows that this change is proportional to $\pm(|A|/m)(1/L^3)$, where the sign is positive for the case in **b** and negative for the case in **c**. If we had N pairs in the box, the change would be proportional to $\pm(|A|/m)n$, where n is the particle density. This simple ‘particle-in-a-box’ picture helps explain why more sophisticated theories show that particle interactions result in an energy term of $4\pi\hbar^2 n/m$.



of new theoretical methods to treat the quantal aspects of the collision in the presence of a large number of internal channels that the collision can follow. The scattering lengths for most of the alkali isotopes are now well known^{17–23}. For example, the Na and ⁸⁷Rb scattering lengths are 55 a_0 and 105 a_0 , respectively, for internal atomic states that have been Bose condensed at low magnetic field, whereas A is negative ($-27 a_0$) for ⁷Li. Scattering lengths can also be tuned by varying a magnetic field (see below).

The low-energy elastic cross-section for the scattering of identical bosonic atoms is $8\pi A^2$, whereas it vanishes for fermions in identical spin states because s -wave scattering is forbidden. The fact that the scattering length can be large compared to the atomic diameter means that the cross-section for elastic collisions can be much larger than that seen in a gas at room temperature where the hard-sphere atomic diameter is the relevant scale. This is particularly important for the success of the evaporative cooling process in alkali gases. It is also possible to have inelastic collisions, which release energy or otherwise result in the formation of untrapped states. Therefore another requirement for successful evaporative cooling is that elastic collision rates greatly exceed inelastic ones. Fortunately this is usually true for a suitable choice of internal state, for which inelastic scattering generally occurs at much smaller distances than the scattering length, and with intrinsically small probabilities.

The description of scattering length, when valid, also makes the calculation of other properties of the gas much easier than they would be if we had to keep track of all of the details of the collision. In the case

of a Bose–Einstein-condensed gas, the average energy of a given particle due to its interaction with the others in the gas is given by $4\pi A\hbar^2 n/m$, where n is the atomic density in the condensate and m is the mass of the atom; this term appears in the Gross–Pitaevskii equation that is solved to obtain the condensate wavefunction (ref. 6; and see review in this issue by Anglin and Ketterle, pages 211–218). The sign of A determines whether this energy term is positive or negative.

Figure 2 indicates why the scattering length determines the interaction energy of low-energy particles. This simple prescription has been shown to produce precise values for the shapes and excitations of condensed gases. It has also been used to make excellent predictions of the time-dependent dynamics that can be observed with condensates, such as four-wave mixing (see review in this issue by Rolston and Phillips, pages 219–224). The fact that the scattering length can be determined independently by the new spectroscopies eliminates the uncertainty in the nature of the interaction between particles that can otherwise plague the theory of many-particle systems. Thus, theoretical calculations on these many-body systems can be done from first principles with accurate, predetermined model parameters.

It has long been known that stable BECs can exist in an infinite homogeneous gas with repulsive interatomic interactions (that is, positive A), but cannot exist in a gas with attractive interactions (negative A). The successful production of a very small ⁷Li condensate is consistent with theoretical predictions that condensation is possible in a small finite-size trap even if A is negative^{24,25}. This is because the quantum zero-point motion in the trap helps to keep the atoms

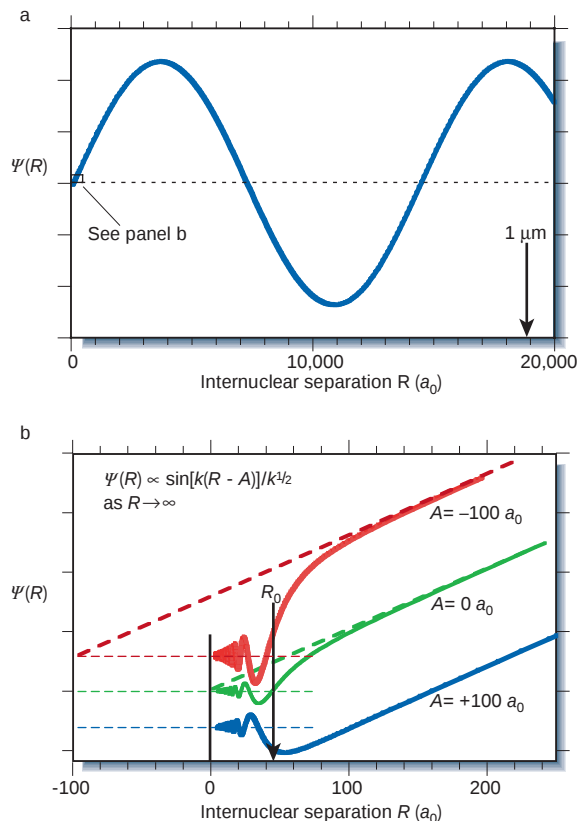
Box 1

Scattering length

The graphs on the right show a typical scattering wavefunction $\Psi(R)$ versus internuclear separation R for the relative motion of two atoms with a ²³Na mass at a collision energy of $E/k_B = 1.4 \mu\text{K}$ and with zero relative angular momentum — that is, an s -wave. The amplitudes of higher states of relative angular momentum, such as p -waves or d -waves, are strongly attenuated at small R owing to long-range centrifugal barriers in the interatomic potential.

Panel a shows the sinusoidal oscillations of the long-range Ψ on the scale of the de Broglie wavelength, which in this case is nearly $1 \mu\text{m}$. The boundary condition requires that $\Psi \rightarrow 0$ as $R \rightarrow 0$. Panel b shows a blow-up of Ψ in the short-range region of strong chemical interactions for three different models of the interaction potential for Na atoms with maximal polarization of electron and nuclear spins. The long-range van der Waals potential, which varies as $-C_6/R^6$, gives rise to a characteristic distance¹ $R_0 = 0.5(mC_6/\hbar^2)^{1/4}$, where m is the atomic mass. For example, $R_0 = 44 a_0$ for ²³Na and $R_0 = 101 a_0$ for ¹³³Cs. At sufficiently low energy, and for $R \gg R_0$, Ψ approaches the asymptotic form $\sin(k(R - A))/k^{1/2}$, where the momentum $p = \hbar k$. When $R < R_0$, Ψ oscillates rapidly on a scale small compared to R_0 , owing to the acceleration of the relative motion by the strong attractive potential. The net effect of the strong interaction is characterized at low E by the single parameter A , called the scattering length. This provides a measure of the phase shift $-kA$ introduced by the collision and is very sensitive to the whole potential for $R < R_0$.

Panel b shows three model examples where A is positive, negative and zero. The long-range sine function acts as if it extrapolates to small R so as to have an energy-independent node at $R = A$ (this would have to be understood as a virtual node for the case of $A < 0$, as $R \geq 0$). This effectively resets the low-energy boundary condition such that the asymptotic sine function seems to vanish at $R = A$, independent of k when k is small. The scattering length is also closely related to the location of the last bound state in the potential. It is large and positive if the last bound state has an energy just below the zero energy threshold, and it is large and



negative if there is a ‘virtual’ bound state just above the threshold. Clearly, the concept of scattering length is useful only if $|A| \ll 1/k = \lambda/2\pi$. It is readily generalized to the case of multiple spin states, as is the situation in H and the alkali atom species.

apart, allowing a small condensate to form, until enough atoms are in the condensate that the attractive forces between the atoms cause it to collapse. Negative scattering lengths also tend to make evaporative cooling more difficult, as was found for ^{85}Rb (ref. 26). This is because the *s*-wave contribution to the cross-section varies with energy in such a way as to pass through a zero at relatively low energy.

So far we have just considered binary collisions in a gas. Three-body interactions can also become important in some cases when the density is sufficiently high. The full solution of the multi-channel, quantal three-body problem remains a major computational challenge. And it will be an important one in future experiments that focus on the interactions between ultracold atoms and molecules.

Collisions in a light field

Collisions in the presence of cooling light usually produce unwelcome effects (that is, heating and loss of trapped atoms) that one tries to avoid when cooling atoms with laser beams¹. The interatomic forces unleashed by coupling to the excited states of a pair of colliding atoms are enormous on the scale of the initial kinetic energies of the ultracold atoms. The resonant dipole interaction energy V_d between a ground-state atom and an excited-state atom already has a magnitude on the order of $V_d/k_B = 1$ mK when the atoms are as far apart as 100 nm. This is the distance at which the colliding atoms are excited by a laser photon in typical cooling experiments. This distance is, of course, large compared to the interatomic distance scale for the 'usual' bound states of the diatomic molecule: these vibrate with typical outer turning points of about $10 a_0$ or less. The strong interatomic force accelerates the atoms together until they either decay by spontaneous emission or undergo a transition to another electronic state. The result is an overall heating of the gas or the ejection of atoms from the trap. Because of this, the laser light is often turned off in order to reach the conditions of low T and high density needed for Bose–Einstein condensation. Fortunately evaporative cooling can be done 'in the dark', or at least without optical frequencies present.

Cold collisions in a light field can also be enormously useful, because the coupling to bound, excited states depends on the precise details of both the ground and excited potential curves and

wavefunctions. This sensitivity can be exploited using photoassociation spectroscopy (refs 27–30; and Fig. 3). Photoassociation is the process where two colliding atoms collectively absorb a photon, forming a bound, excited-state molecule. It can also be driven by more complicated, two-photon or Raman transitions to doubly excited or ground-state molecules (Fig. 4).

Photoassociation spectroscopy is an example of the more general threshold-resonance spectroscopy (Fig. 5), which serves as an exquisitely accurate probe of threshold collisions and the molecular complex made by the colliding atoms. The idea is that a bound state of the dimer complex can be optically or magnetically tuned to be in precise resonance with the colliding atoms, whose energy is constrained to a range on the order of $k_B T$. Because $k_B T$ can be made very small (Fig. 1), the position of the resonance level of the complex can be measured to a precision on the order of its intrinsic width. As a result of the spontaneous emission of light by the excited-state complex, typical one-colour photoassociation spectral widths are limited by the natural line width³¹. These widths are normally several megahertz, that is, comparable to the best Doppler-free conventional molecular spectroscopy. Raman photoassociation can achieve much narrower widths — in fact, Raman resonances with widths on the order of 1 kHz have been observed in ^{87}Rb and ^7Li BECs^{32,33}.

Photoassociation spectroscopy is complementary to the 'usual' bound–bound spectroscopy in molecules, which normally proceeds from deeply bound states and can easily access only the tightly bound excited states that have a good wavefunction overlap with the initial state. Photoassociation, on the other hand, proceeds from slow, colliding atoms with long de Broglie wavelengths and consequently has its best overlaps with the bound states near the dissociation limits in the excited-state potential. This is the region that lends insight into the long-range forces that are interesting in ultracold collisions.

Figure 3c explains how individual spectral features can result from different states of relative angular momentum in the collision. The positions of the features are sensitive to the excited-state potential, whereas the relative intensities of the individual features are very sensitive to the ground-state potential. Detailed theoretical modelling of such spectra leads to accurate and quantitative models

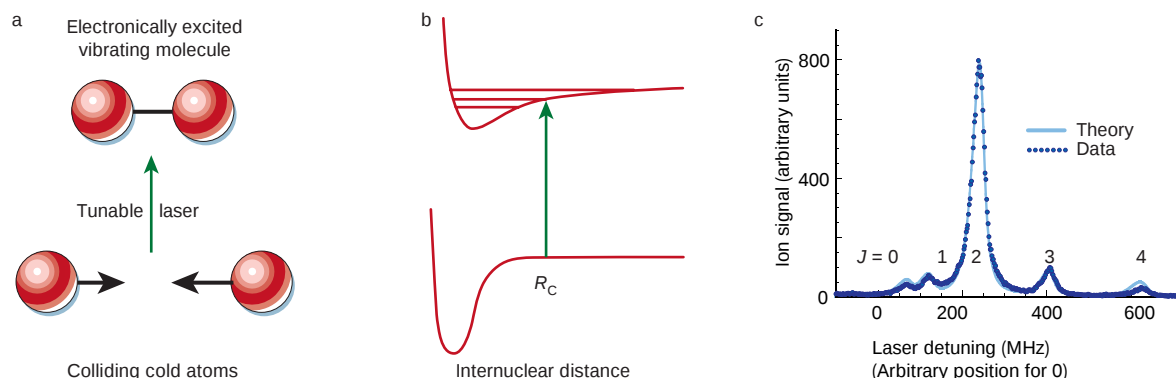


Figure 3 The nature of photoassociation. Two colliding atoms are excited by a tuneable laser to a molecular vibrational state of the atomic dimer molecule. **a**, Schematic illustration of photoassociation. **b**, Electronic potential energy curves for the ground and excited states of the molecule, indicating the presence of quantized vibrational levels in the excited state. The amplitude of the quantum-mechanical excitation process is dominated by contributions near the Condon point R_C , where the R -dependent difference in potential energy curves exactly matches the photon energy $h\nu$. **c**, A sample photoassociation spectrum in a ^{23}Na atom trap³⁰. Photoassociation events are detected either directly, by ionizing the excited-state products, or indirectly, by letting the molecules decay into 'hot' atoms that are no longer held in the trap. In the example shown, the molecular ions made by photo-ionizing the excited level are detected as the excitation laser is tuned over a range of about 700 MHz. This particular

level represents the $v=1$ vibrational level of a Na_2 dimer state with a binding energy of only $E/h = 47$ GHz relative to the separated atom energies. The rotational levels with $J=0$ and 2 are excited from ground state *s*-waves, the $J=1$ and 3 levels are excited from *p*-waves, and the $J=4$ level is excited from a *d*-wave (understanding the specific selection rules requires taking into account the spin structure of the ground and excited states). The *d*-wave feature is weak because the centrifugal barrier for two units of angular momentum is 5 mK, and the trapped atom temperature is only 0.5 mK. The limited penetration of the barrier results in little transition amplitude at the Condon point for this transition. The *p*-wave features are 'accidentally' weak because the *p*-wave ground-state wavefunction has a node near the Condon point. Thus, individual spectral features serve as a probe of the individual components of relative angular momentum of the ground-state wavefunction.

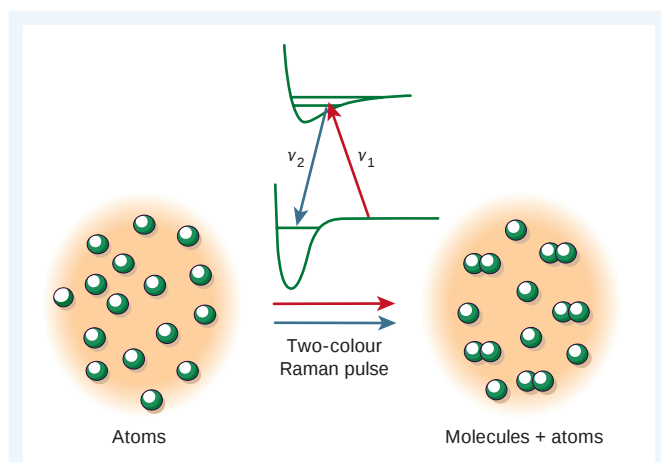


Figure 4 Schematic description of two-colour formation of molecules in their ground electronic state by Raman photoassociation. The potential energy curve shows how two laser fields at frequencies ν_1 and ν_2 can induce a transition from the collisional state of two colliding atoms to a bound vibrational level of the dimer molecule. The difference in photon energies, $h(\nu_1 - \nu_2)$, is tuned near the binding energy of the target molecular level. This is an example of a tuneable threshold-resonance state, which in this case is optically coupled to the colliding atoms. The Franck–Condon principle, which takes into account wavefunction overlap, typically limits favourable coupling to the last few levels in the ground-state potential. Much more accurate values of scattering lengths can be inferred from theoretical models that use the measured binding energies of these last bound states in the potential, just below threshold.

of threshold ground-state scattering that can be applied to other kinds of traps and temperatures. This makes a precise and detailed comparison with the theory of the photoassociation process most fruitful as a way to determine interatomic potentials.

The strong interaction between theory and experiment has been the source of our quantitative understanding of the interatomic forces in the alkalis. It has taken a long time, but we now have comprehensive and quantitative theoretical models of threshold scattering of the collisions of like alkali atom species in their ground electronic states. The models take into account chemical and relativistic spin-dependent interactions and long-range van der Waals dispersion forces, and make use of the best ground-state potential energy curves that are available. They have been refined to account for all available experimental data from photoassociation and other experiments on elastic and inelastic collisions and the last bound states in the potentials.

The characterization of these potentials has permitted ‘applications’ such as the precise measurement of atomic lifetimes by molecular spectroscopy. Precision photoassociation spectroscopy near the dissociation limits can be performed on the excited states of homonuclear diatomic molecules. Accurate theoretical calculations of the appropriate potentials can then be fit to the data to extract accurate values for the model parameters, such as the coefficients that characterize the C_6/R^6 van der Waals and C_3/R^3 resonant dipole contributions to the interatomic potential. The C_3 coefficient is related directly to the lifetime of the asymptotic excited atomic state. Thus molecular spectroscopy leads to the extraction of values of atomic lifetimes that can be more precise than values extracted by more traditional ‘atomic’ methods^{34–36}.

The observation of the effects of the finite speed of light, or ‘retarded potentials’, on the binding energies of spectral lines is another unique feature of the precision spectroscopy of long-range states and of the tremendous accuracy of the theoretical calculations that are now available³⁵. Of course the measurement of scattering lengths for atomic systems, the observation of ‘purely long-range’ states³⁷ (with inner turning points of the vibrational motion at

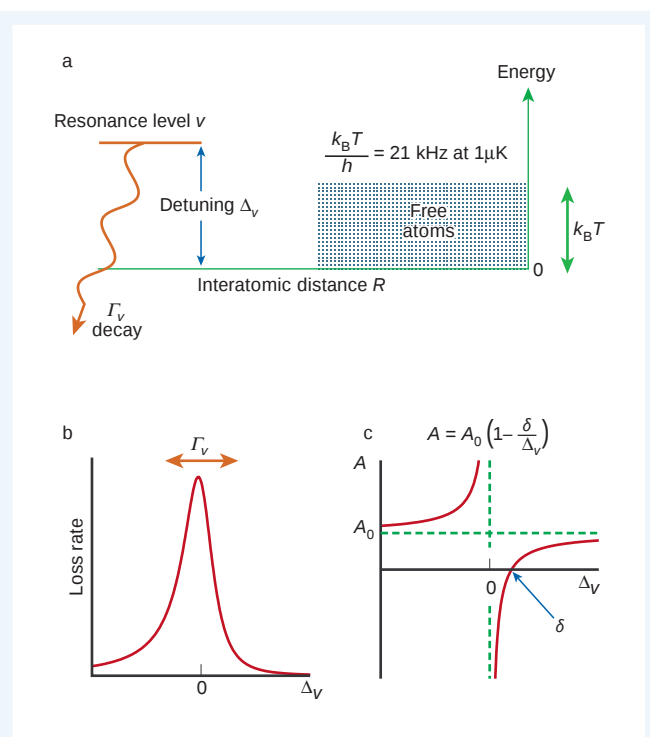


Figure 5 The nature of threshold scattering with a tuneable resonance level.

a, Schematic showing the energetics of a cold collision that results in the formation of a scattering resonance, that is, a quasibound dimer level ν , owing to coupling between the scattering and quasibound states. The coupling may be due to intrinsic chemical or spin-dependent forces, or to optically induced coupling in the case of photoassociation. Resonances that result from the former types of coupling are often called ‘Feshbach’ resonances to distinguish them from photoassociation resonances; but both are treated by the same theoretical formalism and represent different physical ways to realize resonance scattering. The ‘position’ Δ_ν of the resonance relative to the separated atoms can be tuned by varying an external field. This might involve tuning the frequency of the photoassociation laser, as in Fig. 2, or tuning a magnetic field. The resonance typically decays by some process to detectable decay products with a rate Γ_ν/h . One of the most powerful aspects of threshold-resonance spectroscopy with ultracold atoms is that the very narrow distributions of thermal energy permits the energy of the colliding atoms to be characterized precisely. For example, the thermal spread of energy at 1 μK is only $k_B T/h = 21$ kHz. This often is much less than the decay widths of resonance levels, so spectroscopic resolution can be better than even the best Doppler-free optical spectroscopy. This kind of spectroscopy has been carried out both with elastic and inelastic collisions.

b, The characteristic asymmetric shape of the inelastic collision rate versus resonance detuning, which has a width on the order of Γ_ν . **c**, The scattering length for elastic collisions shows a resonance modification as the quasibound state is tuned across the threshold region where Δ_ν is close to zero, and will be large and positive on one side of resonance and large and negative on the other.

distances of several nanometres) and the use of magnetically or optically tuneable resonances for precision spectroscopy all provide examples of the wonderful interplay of theory and experiment in this field. Without precision spectroscopy and the modelling of ultracold gases, the description of low-energy scattering and condensate behaviour would be a truly hit-and-miss affair.

Threshold phenomena for measurement and control

Ultracold atomic gases can be used to study and exploit threshold collision phenomena that have been inaccessible to experimentalists in the past. In addition to photoassociation resonances, a spectacular example of such studies is the observation of magnetically tuneable resonances in the scattering of evaporatively cooled atoms. These resonances lead to a rapid variation in the cross-sections as a function

of the energy of a colliding pair when this energy is nearly coincident with the resonance position. It is crucial for recent studies that the position of the resonance can be tuned using a magnetic field. In such experiments the scattering length, as well as other inelastic ultracold cross-sections, varies in a resonant fashion when the strength of the applied field is varied. The resonant behaviour for the scattering length is described in Fig. 5. The possibility of magnetically tuneable resonances in ultracold collisions was suggested theoretically for collisions between two H atoms³⁸ or two Cs atoms³⁹ and has now been observed in ²³Na and ⁸⁵Rb BECs^{26,40}. This control over the energy of a collision resonance, and thus over the scattering length, can be utilized in several ways. It is now possible to first form a stable condensate with a positive scattering length and then switch the sign of the scattering length. This leads to a spectacular collapse of the now unstable condensate⁷.

The accuracy of measuring the magnetic field at which the resonances are observed translates into accurate bound-state spectroscopy that rivals or exceeds that of photoassociation. An impressive example of ultracold threshold-resonance spectroscopy is shown in Fig. 6 for the collision of two Cs atoms for T on the order of 5 μ K (refs 22,23; and S. Chu, private communication, 2001). Although quantitative theoretical models of threshold Cs scattering have proved to be elusive in the past, magnetic threshold-resonance spectra provide new constraints on the model parameters that solve the problems that were previously encountered. Theory and experiment agree well for several dozen resonances, and molecular quantum numbers can be assigned to each feature in the figure. The surprisingly large scattering lengths for Cs atom collisions mean that the scattering-length approximation fails at remarkably low energy. It cannot be used even at 1 μ K, as $k|A| > 1$ (see Box 1), and full quantum scattering calculations are necessary to model the experiments.

Caesium is especially interesting as it is the basis of the definition of time as agreed by international convention. One of the applications of ultracold atoms is to atomic clocks, and in particular, the new Cs atomic fountain clock (see review in this issue by Udem, Holzwarth and Hänsch, pages 233–237). Atomic collisions are responsible for a density-dependent shift in the frequency determined by the atomic clock. This small shift for Cs is still large enough to be a problem in the establishment of precise time standards. Recent observations have shown that we now have a good enough theory to make accurate determinations of these shifts for a range of proposed clock systems. The model for Cs even predicts that the collision-induced shift in the clock frequency undergoes a surprising variation at very low temperatures (below 1 μ K), even passing through a zero as it changes sign⁴¹. Experiments are underway to test these predictions.

Although threshold resonances have good uses in precision spectroscopy and control of collisions, they may have deleterious side effects, such as collisional losses that limit the lifetime of the cold gas. One-colour photoassociation typically leads to the loss of trapped atoms because of the rapid spontaneous decay of the excited state. Although this loss is advantageous for certain kinds of photoassociation spectroscopy, it means that such resonances are not practical for changing the scattering length. Magnetically tuneable resonances may also lead to three-body losses. This is especially problematical for the experiments with a ²³Na BEC, where strong three-body collisional losses accompanied the tuneable scattering length⁴⁰. Fortunately, such three-body losses are much less prominent in the case of ⁸⁵Rb, where the density is typically kept much lower.

Cold molecules

Molecules cannot easily be laser cooled, because they do not have a closed structure of two levels that can repeatedly scatter light of a specific resonant wavelength. Rather, many rotational and vibrational levels become populated. Translationally cold molecules can be made by photoassociation of two cold atoms. Normal one-colour photoassociation makes excited molecules that rapidly decay by spontaneous emission. A fraction of this incoherent decay process

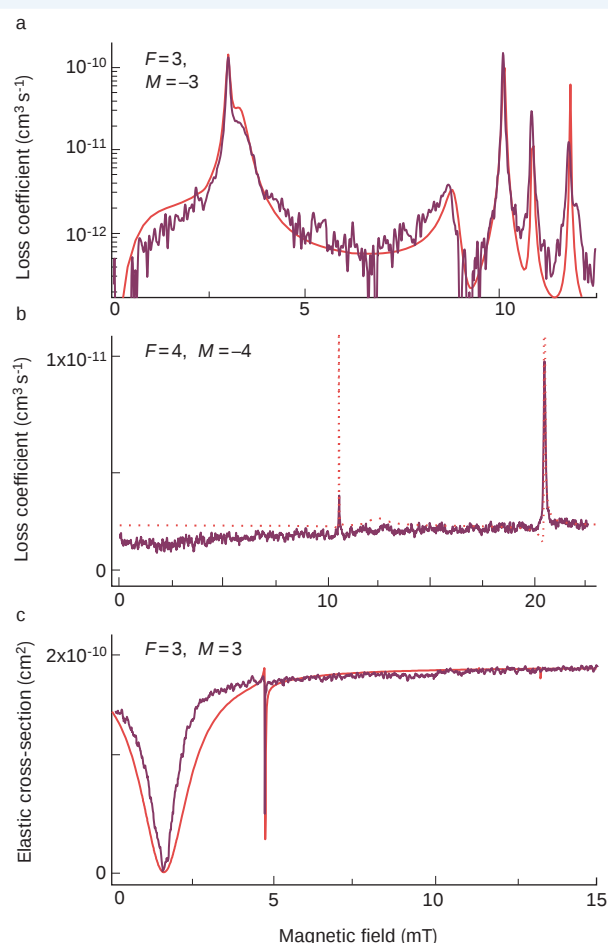


Figure 6 Comparison of experimental and theoretical collisional spectra of cooled and trapped Cs atoms at a temperature of a few microkelvin (refs 22,23; and S. Chu, private communication, 2001). The Cs atoms were confined in an optical trap, so any hyperfine sublevel of the ground electronic state could be held. These levels are labelled by total angular momentum F and its projection M on the magnetic field axis. As the magnetic field is varied, threshold quasibound levels of the Cs_2 dimer molecule are tuned in and out of collisional resonance with the low-energy atoms.

a,b, Collisional rate coefficients for several resonances in inelastic loss collisions of $F=3, M=-3$ atoms (**a**) and $F=4, M=-4$ atoms (**b**). **c**, Elastic scattering cross-section for $F=3, M=3$ atoms. The data curve is normalized to the theoretical one at large magnetic field. The dips in elastic cross-section result from the interference of the resonance scattering amplitude with the background scattering amplitude in the absence of a resonance (at points like those indicated by δ in Fig. 5c).

produces a distribution of ground-state molecular vibrational and rotational levels. In the case of alkali species, these levels could be those of the singlet or triplet electronic states correlating with two ground-state atoms. In fact, cold alkali dimer molecules have been produced in this way^{8–14}. Although the cold molecules might fall out of the atom trap, such molecules can also be trapped using an optical trap at the focus of an intense CO_2 laser beam⁹, or in the lowest triplet state, with a magnetic trap.

Of particular interest would be the ability to make molecules coherently from cold atoms and produce them in specific vibrational–rotational states. One way of doing this is to use a second colour of light to direct the decay of the excited state to a ground state of a specific vibrational–rotational level (Fig. 4). However, the same two laser beams that cause the Raman photoassociation also cause the photodissociation of the molecular level back to separate atoms. It is not possible to build up a large molecular population in a thermal

gas, where the atomic phase space density (Fig. 1) is small compared to unity. One might visualize using short pulses of light to make some molecules, then removing them from the laser excitation region before reapplying the next set of pulses⁴², but it would take a long time to make many molecules in this way in typical magneto-optical traps.

On the other hand, coherent Raman photoassociation of cold atoms might be an effective way to make cold molecules in a BEC⁴³. There is the possibility of a coherent, optically induced conversion between an atomic condensate and a molecular condensate in a matter-wave analogue of second harmonic conversion in nonlinear optics (two atomic field particles combine to produce a single molecular field particle). The gas might even undergo macroscopic oscillations between atomic and molecular forms (analogous to Rabi oscillations)^{44–47}.

In one experiment using coherent Raman photoassociation in a ^{87}Rb atomic condensate, the formation of $^{87}\text{Rb}_2$ molecules that are nearly at rest was inferred from spectroscopic evidence³². The very narrow energy distribution in the condensate and the fine difference-frequency control of the Raman process combine to allow an extremely high-precision measurement of the binding energy of the molecular level relative to the separated atom energies. This measurement will allow theoretical collision models to be calibrated to make an even better prediction of scattering lengths and of the subtle differences in scattering length among various pairs of hyperfine states of ^{87}Rb atoms. These differences are crucial in understanding the properties of multi-component (or spinor) condensates, which are comprised of more than one hyperfine level of the atoms. However, the existence of macroscopic Rabi oscillations of the quantum gas has not been demonstrated experimentally. Indeed, there are theoretical reasons to suppose that the ability to sustain a number of such macroscopic oscillations of matter-wave fields might be limited⁴⁷.

Coherent atomic collisions and entanglement

Atomic collisions have long been viewed as incoherent processes that are responsible for everything from inhomogeneous line broadening to inelastic processes. Today, we realize that coherent collisional processes give rise to the mean field in a BEC. In fact, all collisions are inherently coherent, as the scattering matrix is unitary. In most experiments, researchers want to control collisions to either prevent a process or to create a process that would otherwise not happen. Typically, to use collisions in a coherent manner it is essential to remove or limit the effects that arise from averaging over the thermal motion of the particles.

In the past decade, scientists have been able to effectively cool particles to very near absolute zero and have thus entered a realm where they can control both the motional states of the atoms and their internal states. As a result, collisions become one more tool for creating and controlling atomic systems. Physicists have now manipulated collisions for spectroscopy, BEC dynamics and molecule formation, as discussed above.

Far more exotic ideas are now being pursued. Several groups have proposed using atom-atom interactions or collisions to create entanglement between atoms^{48–50}. Such controlled entanglement could be used for building a quantum computer, for teleportation of quantum information, or for improved precision measurements. The basic concept involves loading single atoms into individual microtraps. These atoms will then be brought together to interact or collide in a conditional manner so that the ‘fate of atom 1’ becomes intertwined with the ‘fate of atom 2’. In other words, they are entangled. Once these atoms have their fates mutually tied together the atoms will then be separated.

Two entangled atoms can be additionally entangled with other atoms to make more massive entanglement; this has been shown in elegant experiments with ions⁵¹. Controlled entanglement could be used for engineering exotic quantum states for better interferometers, improved gyroscopes and precision measurement. Controlled entanglement is a fundamental resource needed for quantum information

and quantum computing (see review in this issue by Monroe, pages 238–246). Here the object is to entangle numerous atoms in the process of performing computation by conditionally interacting individual pairs of atoms, either sequentially or in parallel.

Performing such entanglement operations requires individual atom control, which might be achieved by loading atoms into optical lattices or optical/magnetic microtraps formed above the surfaces of microfabricated ‘atom chips’. The traps would act both to localize atoms in unique sites, perhaps to a few tens of nanometres, and to provide control over the two-atom entangling operations. These individually trapped atoms could act as quantum bits, or qubits, by taking advantage of their internal structure — hyperfine states or long-lived resonant states — or by using the motional states that result from trapping the atoms.

Future challenges

Ultracold collisions have had a significant experimental and theoretical impact across a broad spectrum of atomic physics. The generation of cold molecules is poised to further alter our view of the low-energy world. Unprecedented control has been gained over collisions at these temperatures and our investigations will continue to unveil a host of wonderful new phenomena, many of which are uniquely quantum mechanical in nature. One remarkable aspect of the study of ultracold collisions is the ability to characterize the atom-atom interaction by a single parameter, the scattering length. This has led to some of the first tests of many-body physics with no adjustable or phenomenological parameters. The basic control over collisions that is attainable in these systems has led to better precision measurements and will lead to even further improvements. Ultimately, this precise level of control of individual atoms or pairs of atoms may lead to the development of physical devices to be used for quantum information processing. The entanglement of the quantum states of atoms that result from their collisional interactions may make possible the creation of correlated beams of atoms^{52,53}.

We have concentrated here on trapped atoms, and the size of the trap has been assumed to be much larger than any size that characterizes the interactions in the collision (for example, the scattering length). Therefore, a free-space description of the atomic collisions has been adequate. But if the trap is made very small in one or more of its dimensions, the confinement can have a profound effect on the collision⁵⁴. New experimental and theoretical challenges will include the orchestration of collisions in constrained geometries. Tight traps can constrain the collisions to occur in a single cell or along a single axis or in a single plane. These situations correspond to zero-, one- and two-dimensional scattering, respectively, and will modify the free-space collision properties. Additionally, such reduced dimensionality can lead to interesting new quantum dynamics and phase transitions⁵⁵.

Several groups are now making cold molecules by assembling them from already cold atoms. In addition, cold molecules have recently been produced and trapped without using laser cooling. One method, applied to the CaH diatomic molecule, used a helium buffer gas to cool this paramagnetic species to a temperature low enough that it could be held in a magnetic trap¹⁵. Another group constructed a pulsed electric-field decelerator to slow a beam of polar ammonia molecules and then load them into an electric quadrupole trap¹⁶. These new methods could be applied to a large number of different molecular species (paramagnetic and dipolar, respectively). This could extend precision measurements and quantum-level control to atom-molecule and molecule-molecule collisions, the study of which is just beginning. Characterization of the potentials for these systems is clearly required for making progress. Trapped molecules may also offer some real advantages in the elusive search for the dipole moment of the electron. Evaporative cooling should work for molecules, if inelastic collisions do not lead to heating and loss, and perhaps could result in the formation of molecular condensates.

The extension of these achievements to more complicated molecules, should that prove to be possible, may lead to new ways to probe

collision complexes and to the control of all degrees of freedom of molecules undergoing chemical reactions. The ability to control individual atoms or molecules might be used for the nanofabrication of 'designer molecules' or the manipulation of individual biomolecules.

We have already seen our dreams of controlling interactions on the quantum level come true, and the exquisite nature of this control has proved remarkable. These achievements have come from experimental and theoretical developments that have been a joy to be involved in, and their impact on new physics, chemistry and quantum computation has only just begun. □

- Weiner, J., Zilio, S., Bagnato, V. S. & Julienne, P. S. Experiments and theory in cold and ultracold collisions. *Rev. Mod. Phys.* **71**, 1–85 (1999).
- Cohen-Tannoudji, C. Manipulating atoms with photons. *Rev. Mod. Phys.* **70**, 707–719 (1997).
- Chu, S. The manipulation of neutral particles. *Rev. Mod. Phys.* **70**, 685–706 (1997).
- Phillips, W. D. Laser cooling and trapping of neutral atoms. *Rev. Mod. Phys.* **70**, 721–741 (1997).
- Ketterle, W. & Van Druten, N. J. Evaporative cooling of trapped atoms. *Adv. At. Mol. Opt. Phys.* **37**, 181–236 (1996).
- Dalfovo, F., Giorgini, S., Pitaevskii, L. P. & Stringari, S. Theory of Bose-Einstein condensation in trapped gases. *Rev. Mod. Phys.* **71**, 463–512 (1999).
- Roberts, J. L. *et al.* Controlled collapse of a Bose-Einstein condensate. *Phys. Rev. Lett.* **86**, 4211–4214 (2001).
- Fioretti, A. *et al.* Formation of cold Cs₂ molecules through photoassociation. *Phys. Rev. Lett.* **80**, 4402–4405 (1998).
- Takekoshi, T., Patterson, B. M. & Knize, R. J. Observation of cold ground-state cesium molecules produced in a magneto-optical trap. *Phys. Rev. A* **59**, R5–R7 (1999).
- Nikolov, A. N. *et al.* Observation of ultracold ground-state potassium molecules. *Phys. Rev. Lett.* **82**, 703–706 (1999).
- Nikolov, A. N. *et al.* Efficient production of ground-state potassium molecules at sub-mK temperatures by two-step photoassociation. *Phys. Rev. A* **64**, 246–249 (2000).
- Band, Y. B. & Julienne, P. S. Ultracold molecule production by laser cooled atom photoassociation. *Phys. Rev. A* **51**, R4317–R4320 (1995).
- Gabbanini, C., Fioretti, A., Lucchesini, A., Gozzini, S. & Mazzoni, M. Cold rubidium molecules formed in a magneto-optical trap. *Phys. Rev. Lett.* **84**, 2814–2817 (2000).
- Fatemi, F. K., Jones, K. M., Lett, P. D. & Tiesinga, E. Ultracold ground state molecule production in sodium. *Phys. Rev. A* (in the press).
- Weinstein, J. D., deCarvalho, R., Guillet, T., Friedrich, B. & Doyle, J. M. Magnetic trapping of calcium monohydride molecules at millikelvin temperature. *Nature* **395**, 148–150 (1998).
- Bethlem, H. L. *et al.* Electrostatic trapping of ammonia molecules. *Nature* **406**, 491–494 (2000).
- Abraham, E. R. I., McAlexander, W. I., Gerton, J. M. & Hulet, R. G. Triplet s-wave resonance in ⁶Li collisions and scattering lengths of ⁶Li and ⁷Li. *Phys. Rev. A* **55**, R3299–R3302 (1997).
- van Abeelen, F. A. & Verhaar, B. J. Determination of collisional properties of cold Na atoms from analysis of bound-state photoassociation and Feshbach resonance field data. *Phys. Rev. A* **59**, 578–584 (1999).
- Crubellier, A. *et al.* Simple determination of Na₂ scattering lengths using observed bound levels at the ground state asymptote. *Eur. Phys. J.* **6**, 211–220 (1999).
- Bohn, J. L. *et al.* Collisional properties of ultracold potassium: consequences for degenerate Bose and Fermi gases. *Phys. Rev. A* **59**, 3660–3664 (1999).
- Burke, J. P. Jr, Bohn, J. L., Esry, B. D. & Greene, C. H., Prospects for mixed-isotope Bose-Einstein condensates in rubidium. *Phys. Rev. Lett.* **80**, 2097–3000 (1998).
- Chin, C., Vuletic, V., Kerman, A. J. & Chu, S. High resolution Feshbach spectroscopy of cesium. *Phys. Rev. Lett.* **85**, 2717–2720 (2000).
- Leo, P. J., Williams, C. J. & Julienne, P. S. The collision properties of ultracold ¹³³Cs atoms. *Phys. Rev. Lett.* **85**, 2721–2724 (2000).
- Dodd, R. J. *et al.* Role of attractive interactions on Bose-Einstein condensation. *Phys. Rev. A* **54**, 661–664 (1996).
- Bradley, C. C., Sackett, C. A. & Hulet, R. G. Bose-Einstein condensation of lithium: observation of limited condensate number. *Phys. Rev. Lett.* **78**, 985–989 (1997).
- Cornish, S. L., Claussen, N. R., Roberts, J. L., Cornell, E. A. & Wieman, C. E. Stable ⁸⁵Rb Bose-Einstein condensates with widely tuneable interactions. *Phys. Rev. Lett.* **85**, 1795–1798 (2000).
- Lett, P. D., Julienne, P. S. & Phillips, W. D. Photoassociative spectroscopy of laser-cooled atoms. *Annu. Rev. Phys. Chem.* **46**, 423–452 (1995).
- Heinzen, D. J. in *Atomic Physics* Vol. 14 (eds Wineland, D., Wieman, C. & Smith, S. J.) 369–388 (AIP Press, New York, 1995).
- Stwalley, W. C. & Wang, H. Photoassociation of ultracold atoms: a new spectroscopic technique. *J. Mol. Spect.* **195**, 194–228 (1999).
- Tiesinga, E. *et al.* A spectroscopic determination of scattering lengths for sodium atom collisions. *J. Res. Natl Inst. Stand. Tech.* **101**, 505–520 (1996).
- Napolitano, R., Weiner, J., Williams, C. J. & Julienne, P. S. Line shapes of high resolution photoassociation spectra of optically cooled atoms. *Phys. Rev. Lett.* **73**, 1352–1355 (1994).
- Wynar, R., Freeland, R. S., Han, D. J., Ryu, C. & Heinzen, D. J. Molecules in a Bose-Einstein condensate. *Science* **287**, 1016 (2000).
- Gerton, J. M., Strekalov, D., Prodan, I. & Hulet, R. G. Direct observation of growth and collapse of a Bose-Einstein condensate with attractive interactions. *Nature* **408**, 692–695 (2000).
- McAlexander, W. I., Abraham, E. R. I. & Hulet, R. G. Radiative lifetime of the 2P state of lithium. *Phys. Rev. A* **54**, R5–R8 (1996).
- Jones, K. *et al.* Measurement of the atomic Na(3P) lifetime and of retardation in the interaction between two atoms bound in a molecule. *Europhys. Lett.* **35**, 85–90 (1996).
- Wang, H. *et al.* Precise determination of the dipole matrix element and radiative lifetime of the ³⁹K 4p state by photoassociative spectroscopy. *Phys. Rev. A* **55**, R1569–R1572 (1997).
- Stwalley, W. C., Uang, Y.-H. & Pichler, G. Pure long-range molecules. *Phys. Rev. Lett.* **41**, 1164–1167 (1978).
- Stwalley, W. C. Stability of spin-aligned hydrogen at low temperatures and high magnetic fields: new field-dependent scattering resonances and predissociations. *Phys. Rev. Lett.* **37**, 1628–1631 (1981).
- Tiesinga, E., Verhaar, B. J. & Stoof, H. T. C. Threshold and resonance phenomena in ultracold ground-state collisions. *Phys. Rev. A* **47**, 4114 (1993).
- Inouye, S. *et al.* Observation of Feshbach resonances in a Bose-Einstein condensate. *Nature* **392**, 151 (1998).
- Leo, P., Julienne, P. S., Mies, F. H. & Williams, C. J. Collisional frequency shifts in ¹³³Cs fountain clocks. *Phys. Rev. Lett.* **86**, 3743–3746 (2001).
- Vardi, A., Abrashkevich, D., Frishman, E. & Shapiro, M. Theory of radiative recombination with strong laser pulses and the formation of ultracold molecules via stimulated photo-recombination of cold atoms. *J. Chem. Phys.* **107**, 6166–6174 (1997).
- Julienne, P. S., Burnett, K., Band, Y. B. & Stwalley, W. C. Stimulated Raman molecule production in Bose-Einstein condensates. *Phys. Rev. A* **58**, R797–R800 (1998).
- Timmermans, E., Tommasini, P., Cote, R., Hussein, M. & Kerman, A. Rarefied liquid properties of hybrid atomic-molecular Bose-Einstein condensates. *Phys. Rev. Lett.* **83**, 2691–2694 (1999).
- Javanainen, J. & Mackie, M. Coherent photoassociation of a Bose-Einstein condensate. *Phys. Rev. A* **59**, R3186–R3189 (1999).
- Heinzen, D. J., Wynar, R., Drummond, P. D. & Kheruntsyan, K. V. Superchemistry: dynamics of coupled atomic and molecular Bose-Einstein condensates. *Phys. Rev. Lett.* **84**, 5029–5033 (2000).
- Holland, M., Park, J. & Walser, R. Formation of pairing fields in resonantly coupled atomic and molecular Bose-Einstein condensates. *Phys. Rev. Lett.* **86**, 1915–1918 (2001).
- Brennen, G. K., Caves, C. M., Jessen, P. S. & Deutsch, I. H. Quantum logical gates in optical lattices. *Phys. Rev. Lett.* **82**, 1060–1063 (1999).
- Jaksch, D., Briegel, H.-J., Cirac, J. I., Gardiner, C. W. & Zoller, P. Entanglement of atoms via cold controlled collisions. *Phys. Rev. Lett.* **82**, 1975–1978 (1999).
- Jaksch, D. *et al.* Fast quantum gates for neutral atoms. *Phys. Rev. Lett.* **85**, 2208–2211 (2000).
- Sackett, C. A. *et al.* Experimental entanglement of four particles. *Nature* **404**, 256–259 (2000).
- Pu, H. & Meystre, P. Creating macroscopic atomic Einstein-Podolsky-Rosen states from Bose-Einstein condensates. *Phys. Rev. Lett.* **85**, 3987–3990 (2000).
- Duan, L.-M., Sorensen, A., Cirac, J. I. & Zoller, P. Squeezing and entanglement of atomic beams. *Phys. Rev. Lett.* **85**, 3991–3994 (2000).
- Tiesinga, E., Williams, C. J., Mies, F. H. & Julienne, P. S. Interacting atoms under strong quantum confinement. *Phys. Rev. A* **61**, 063416–1–063416–8 (2000).
- Olshanii, M. Atomic scattering in the presence of an external confinement and a gas of impenetrable bosons. *Phys. Rev. Lett.* **81**, 938–941 (1998).
- Jessen, P. S. & Deutsch, I. H. Optical lattices. *Adv. At. Mol. Opt. Phys.* **36**, 95–137 (1996).

Acknowledgements

This work was supported by the US Office of Naval Research, the EPSRC and the EU.

Optical frequency metrology

Th. Udem, R. Holzwarth & T. W. Hänsch

Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Strasse 1, 85748 Garching, Germany (e-mail: thomas.udem@mpq.mpg.de)

Extremely narrow optical resonances in cold atoms or single trapped ions can be measured with high resolution. A laser locked to such a narrow optical resonance could serve as a highly stable oscillator for an all-optical atomic clock. However, until recently there was no reliable clockwork mechanism that could count optical frequencies of hundreds of terahertz. Techniques using femtosecond-laser frequency combs, developed within the past few years, have solved this problem. The ability to count optical oscillations of more than 10^{15} cycles per second facilitates high-precision optical spectroscopy, and has led to the construction of an all-optical atomic clock that is expected eventually to outperform today's state-of-the-art caesium clocks.

For more than a century, precise spectroscopy of atoms and molecules has been crucial in the discovery of the laws of quantum physics, in the determination of fundamental constants, and in the realization of standards for time, frequency and length. The advent of highly monochromatic, tunable lasers and techniques for nonlinear Doppler-free spectroscopy in the early 1970s had a marked impact on the field of precision spectroscopy^{1,2}. Today, we are able to observe extremely narrow optical resonances in cold atoms or single trapped ions, with resolutions ranging from 10^{-13} to 10^{-15} , so that it might ultimately become possible to measure the line centre of such a resonance to a few parts in 10^{18} . Laboratory experiments searching for slow changes of fundamental constants would then reach unprecedented sensitivity. An all-optical atomic clock³ based on such resonances could satisfy the growing demands of optical frequency metrology, fibre-optical telecommunication or navigation. However, until recently, the lack of an optical frequency counter to act as the clockwork mechanism prevented the construction of such a device.

Most spectroscopic experiments still rely on a measurement of optical wavelengths rather than frequencies. Unavoidable distortions in the geometric wavefront have so far made it impossible to exceed an accuracy of a few parts in 10^{10} with a laboratory-sized wavelength interferometer. To obtain highly accurate results it is necessary to measure the frequency of light rather than its wavelength. This is because time can be measured much more precisely than any other physical quantity, and counting the number of cycles in a second is as accurate as the clock that is used to determine the duration of the second. Because the value of the speed of light was defined in 1983 to be precisely $299,792,458 \text{ m s}^{-1}$ in vacuum, the conversion between frequency and wavelength can be done without loss in accuracy. But the potential for high accuracy of frequency measurement has been restricted mainly to radio frequencies (up to 100 GHz) for many years, and it has been extremely difficult to extend it to the domain of rapid optical oscillations.

The early approach to this problem made use of harmonic frequency chains. Such chains start with a caesium atomic clock that operates, by definition of the SI second, at 9,192,631,770 Hz (equivalent to the ground-state hyperfine splitting of the caesium atom). This clock defines the frequency at the lower end of the chain from which higher harmonics in nonlinear diode mixers, crystals and other nonlinear devices are generated⁴⁻⁸. Phase-locked transfer

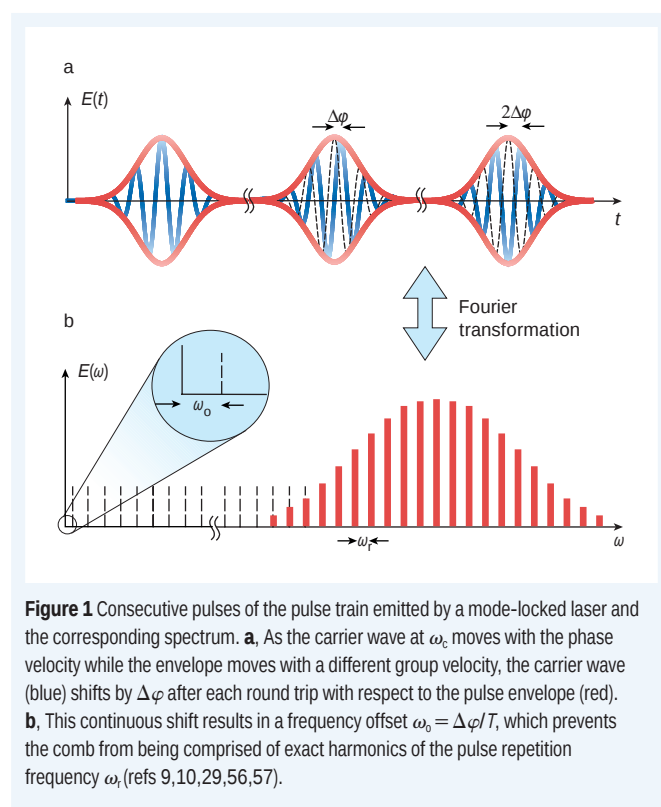
oscillators are needed after each step, so that a chain traversing a vast region of the electromagnetic spectrum becomes highly complex, large and delicate, and requires substantial resources and considerable efforts to build and operate. This is the reason why only a few harmonic chains have ever been constructed. Another significant drawback of this approach is that the chains are designed to measure just one single optical frequency.

None of the problems associated with harmonic frequency chains has been solved since the first chain was constructed about 30 years ago⁴. In 1998 our laboratory introduced a revolutionary new approach that vastly simplifies optical frequency measurements^{9,10}. We showed that the modes of a mode-locked femtosecond laser can be used as a precise ruler in frequency space, as they form a series of frequency spikes called a frequency comb^{11,12}. This work has now culminated in a compact and reliable all-solid-state optical frequency synthesizer that requires only a single mode-locked laser, but is nevertheless capable of measuring essentially any optical frequency¹³⁻¹⁷. As a universal optical frequency-comb synthesizer it provides the missing link between optical and microwave frequencies. For the first time, small-scale spectroscopy laboratories can measure or synthesize any optical frequency with extreme precision. Femtosecond frequency-comb techniques have since started to gain widespread use, with precision measurements in Rb (ref. 13), Ca (refs 18,19), CH_4 (ref. 20), H (refs 15,16), Hg^+ (ref. 18), I_2 (refs 14,21-24), Sr^+ (refs 24,25), Yb^+ (refs 24,26) and In^+ (ref. 27).

The same femtosecond frequency-comb techniques are also opening new frontiers in ultrafast physics. Control of the phase evolution of few-cycle light pulses^{9,17,28-30} provides a powerful new tool for the study of highly nonlinear phenomena that depend on the phase of the carrier wave relative to the pulse envelope, such as above-threshold ionization³¹, strong-field photoemission or the generation of soft-X-ray attosecond pulses by high-order harmonic generation³².

Femtosecond light pulses

More than 20 years ago, the frequency comb of a mode-locked picosecond dye laser was first used as a ruler in frequency space to measure the sodium $4d$ fine-structure splitting³³. This route was pursued further in the 1980s (refs 34,35), but the attainable bandwidths were never large enough to make it a widespread technique for optical frequency metrology. Broadband femtosecond Ti:sapphire lasers have existed since the beginning of the



1990s, and we have shown conclusively that such lasers can be crucial in this field^{11,12}.

To understand the mode structure of a femtosecond frequency comb and the techniques applied for its stabilization, consider the idealized case of a pulse circulating in a laser cavity with length L with a carrier frequency ω_c , as shown in Fig. 1. The output of this laser is a sequence of pulses that are essentially copies of the same pulse separated by the round-trip time $T = v_g/2L$, where v_g is the cavity's mean group velocity defined by the round-trip time and the cavity length. But the pulses are not quite identical. This is because the pulse envelope $A(t)$ propagates with v_g , whereas the carrier wave travels with its phase velocity. As a result, the carrier shifts with respect to the pulse envelope after each round trip by a phase angle $\Delta\phi$ (Fig. 1). Unlike the envelope function, which provides us with a more rigorous definition of the pulse repetition time $T = \omega_r^{-1}$ by demanding $A(t) = A(t - T)$, the electric field is, in general, not expected to be periodic in time. If the periodicity of the envelope function is assumed, the electric field at a given place outside the laser resonator can be written as

$$E(t) = \text{Re}(A(t) \exp(-i\omega_c t)) = \text{Re}(\sum_n A_n \exp(-i(\omega_c + n\omega_r)t)) \quad (1)$$

where A_n are Fourier components of $A(t)$. This equation shows that, under the assumption of a periodic pulse envelope, the resulting spectrum consists of a comb of laser modes that are separated by the pulse repetition frequency ω_r .

Because ω_c is not necessarily an integer multiple of ω_r , the modes are shifted from being exact harmonics of the pulse repetition frequency by an offset that can be chosen to obey $\omega_o < \omega_r$ simply by renumbering the modes

$$\omega_n = n\omega_r + \omega_o \quad (2)$$

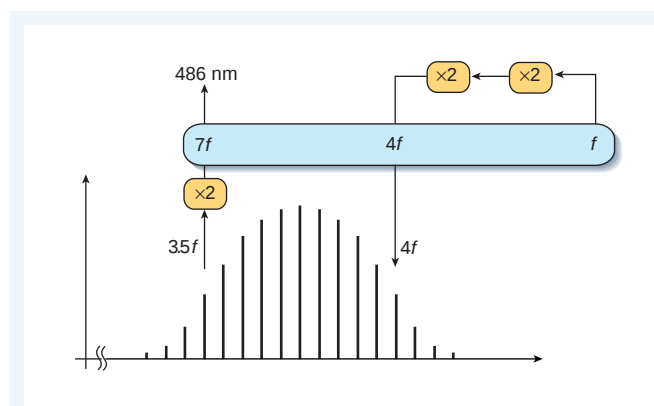
with a large ($\approx 10^5$ – 10^6) integer n . This equation maps two radio frequencies ω_r and ω_o onto the optical frequencies ω_n . Although ω_r is readily measurable, and usually lies between a few tens of megahertz and a few gigahertz depending on the length of the laser resonator, ω_o

is not easy to access unless the frequency comb contains more than an optical octave. The intuitive picture given here can even cope with a frequency chirp, that is, a carrier frequency that varies across the pulse. In this case the envelope function becomes complex in value and the comb structure derived above remains valid provided the chirp is the same for all the pulses. Under this assumption, which is reasonable for a stationary pulse train, $A(t)$ remains a periodic function.

The main presumption of equation (2) is that the mode spacing is the same across the frequency comb and that it equals the pulse repetition rate. To be able to use this property for high-precision measurements, one has to make sure that the simple, intuitive picture given above is correct with a high degree of accuracy. Initially, the crucial question was whether pulse-to-pulse phase fluctuations or the large multiplication factor n might destroy the coherence³⁶ or prevent it from being a comb all together. We first verified that there is indeed a comb of continuous-wave laser frequencies by overlapping the beam from the femtosecond laser with a single-mode diode laser on a photo-detector to observe beat signals between the modes and the continuous-wave laser. To test the mode spacing we have used two laser diodes that were phase locked to distant modes of the comb and used simultaneously as an input for a so-called optical frequency interval divider (OFID)^{37,38}. An OFID locks three laser frequencies f_1 , f_2 and f_3 to obey the relation $2f_3 = f_1 + f_2$ in a phase-coherent manner. This could be used, for example, to produce an output (f_3) at the precise midpoint of the frequency interval given by two input laser diodes (f_1 and f_2). If the modes are distributed equally on the frequency axis, and if there is an odd number of modes between the two OFID inputs, then one would expect the OFID output to coincide with another mode of the comb. After averaging all our data we concluded that there was no deviation from the perfect regular grid larger than 3 parts in 10^{17} (ref. 11). This experiment was key to the development that followed.

Improving the versatility of frequency measurements

At this point, the frequency comb generated by our mode-locked Ti:sapphire laser could bridge roughly 45 THz. The fourth harmonic of a methane-stabilized He–Ne laser that operated with high accuracy was used as an absolute frequency reference for the femtosecond comb, which allowed us to access new targets. These included the transition frequency of the caesium D_1 line¹², which is needed for a new determination of the fine-structure constant α (refs 39,40). And a sharp resonance in a single trapped indium ion²⁷ may eventually serve as a precise reference for an optical clock⁴¹. With the previous measurement schemes it would have been very difficult to measure



the frequency gap from the fourth harmonic of the He–Ne reference to these target frequencies, but with the frequency comb this was a straightforward task.

Direct radio frequency–optical frequency link

Combs used as rulers in frequency space can measure large frequency gaps between a precisely known optical reference and an unknown frequency. But to reference the comb directly to a precisely known radio frequency an absolute optical frequency could be mapped onto a frequency difference between harmonics or sub-harmonics of the same laser. In the simplest case this would be the interval between an optical frequency f and its own second harmonic $2f$ (ref. 14). The measurement of the large frequency gap $f = 2f - f$ by a chain of OFIDs or a femtosecond laser has been proposed in refs 37 and 9, respectively. But other intervals can be used as well, as illustrated in Fig. 2, which describes the first experiment of this kind^{15,42}. The experimental set-up still uses the somewhat awkward quadrupling stage to create the fourth harmonic $4f$ of the methane-stabilized He–Ne laser. The frequency comb is then used to cross over from $4f$ to $\sim 3.5f$, and after doubling, a frequency in the vicinity of $7f$ is obtained. The loop is then closed by using an OFID stage that fixes f such that frequency ratios $7f:4f:f$ are precisely fulfilled ($2 \times 4f = 7f + f$). The easiest way to contemplate this is to note that the frequency comb locks $4f - 3.5f = 0.5f$ (44 THz) to an integer multiple of the caesium clock that controlled the mode spacing. By controlling $0.5f$ with a caesium atomic clock, we also know the frequency f and every other frequency in the set-up, including every mode of the comb, with the precision of the caesium clock. Thus the stabilization on methane is replaced by locking all frequencies directly to the caesium clock.

We have used the 486-nm output of this set-up (with some modifications not shown) to measure the absolute frequency of the hydrogen 1S–2S two-photon resonance, which occurs at the fourth harmonic of this wavelength. The mode spacing was controlled by a caesium atomic fountain clock constructed by André Clairon and co-workers at the Laboratoire Primaire du Temps et des Fréquences (LPTF, France). The outcome was one of the most precise measurements of an optical frequency so far recorded¹⁶, providing a transition frequency value as accurate as 1.9 parts in 10^{14} , limited by the reproducibility of the hydrogen spectrometer. Together with other transition frequencies in hydrogen, our value for the 1S–2S resonance has resulted in much improved values for the Rydberg constant, which is now one of the most accurately known fundamental constants, and the Lamb shift of the 1S state, which now provides one of the most stringent tests of quantum electrodynamics⁷.

Increasing the bandwidth of frequency combs

The first absolute measurement of an optical frequency with a femtosecond frequency comb has inspired further rapid advances in the art of frequency metrology. Spectral broadening resulting from self-phase modulation via the intensity-dependent index of refraction in an optical fibre was used to increase the widths of the frequency combs. Even though the dispersion inside the fibre changes the shape of the pulses significantly, it does so in the same way for all pulses. As the above arguments that led to the regular-spaced frequency comb assumed only the periodicity of the envelope function, they should also apply to whatever comes out of the fibre. With newly invented micro-structured photonic crystal fibres (PCF)^{43–45}, this spectral broadening is so pronounced that the resulting frequency comb can span more than an optical octave, even with the moderate output power from the laser oscillator. In collaboration with Phillip Russell, Jonathan Knight and William Wadsworth from the University of Bath, we have used such fibres to construct a much more compact version of the set-up (Fig. 3).

Alternative approaches to produce large-bandwidth frequency combs have also emerged. Short fibre tapers are now used for this purpose⁴⁶; these are manufactured from standard telecom fibres by drawing them in a flame until their diameter shrinks to about 2 μm .

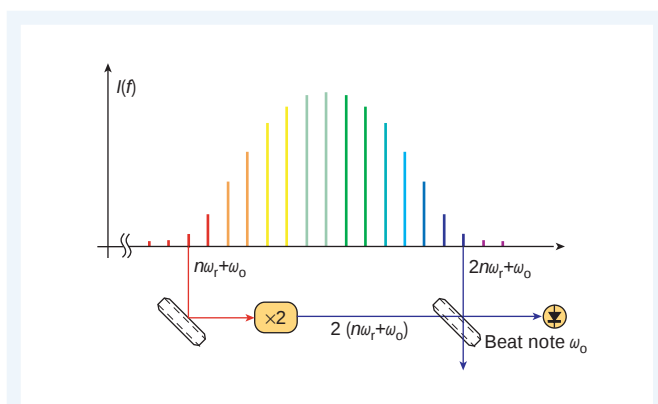


Figure 3 The principle of the single-laser optical synthesizer. A mode with the mode number n at the red wing of the comb and whose frequency is given according to equation (2) by $\omega_n = n\omega_l + \omega_o$ is frequency doubled in a nonlinear crystal. If the frequency comb covers a full optical octave, a mode with the number $2n$ should oscillate simultaneously at $\omega_{2n} = 2n\omega_l + \omega_o$. The beat note between the frequency-doubled mode and the mode at $2n$ yields the offset frequency $2(n\omega_l + \omega_o) - (2n\omega_l + \omega_o) = \omega_o$.

Just like in a PCF, the small core, which consists of the whole fibre, enhances the nonlinear interaction while the evanescent wave around the taper causes a smaller group-velocity dispersion. Another method uses the high intensity inside the laser cavity to efficiently broaden the spectrum⁴⁷.

The compact optical frequency synthesizer

With an octave-spanning frequency comb (Fig. 3), we can directly access the interval between an optical frequency f and its second harmonic $2f$, where f could be the 1064-nm line of a Nd:YAG laser. This greatly simplifies the experimental set-up shown in Fig. 2 and has first been reported by John Hall's group at the Joint Institute for Laboratory Astrophysics (JILA) in Boulder, Colorado, and by our group at the Max Planck Institute for Quantum Optics (MPQ) in Garching, near Munich^{13–15,17}. The JILA group has also shown that

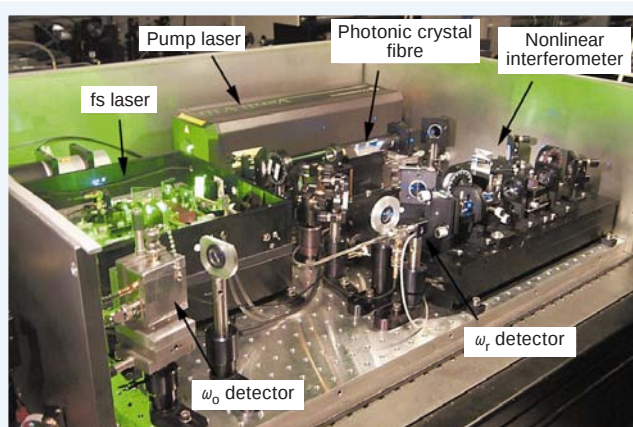


Figure 4 An optical synthesizer in a box. The device consists of one femtosecond laser powered by a green pump laser (Verdi model, Coherent), a photonic crystal fibre and a nonlinear interferometer (as illustrated in Fig. 3). It occupies only 1 m² on an optical bench with the potential for further miniaturization. The synthesizer is capable of linking a 10-MHz radio-frequency reference phase coherently in one step with the optical region, and provides a reference-frequency grid across much of the visible and infrared spectrum with comb lines that are separated by 625 MHz. This makes it an ideal laboratory tool for precision spectroscopy. The optical frequency synthesizer is now commercially available from Menlo Systems GmbH (www.menlosystems.com), co-founded by two of the authors (R.H. and T.W.H.).

the additional Nd:YAG laser was not necessary if the red part of the frequency comb itself is to be frequency doubled.

With this set-up, the two radio frequencies that enter equation (1), ω_r and ω_o , are determined. The pulse repetition rate ω_r is simply measured with a photo-detector anywhere in the beam and ω_o , which is smaller in value than ω_r , is derived as explained in Fig. 3. To obtain a stable comb for absolute optical frequency measurements, it is advantageous to phase lock both ω_r and ω_o to a precise radio-frequency reference such as a caesium atomic clock or Global Positioning System-controlled quartz oscillator. The mode number n may be determined by a coarse measurement of the mode in question, for example with a wavemeter. As soon as n , ω_r and ω_o are known, every frequency contained in the comb is known with roughly the same precision. Figure 4 shows the actual apparatus used in our laboratory.

To check the integrity of the broad frequency comb and evaluate the overall performance of the optical synthesizer we have compared it with the previous $7f:4f:f$ variant. To compare the two we use its output at $4f$ (354 THz) and a mode at 354 THz of the $f:2f$ optical synthesizer. After averaging all data we obtained a mean deviation from the expected beat frequency of 71 ± 179 mHz at 353 THz. This corresponds to a relative uncertainty of 5.1×10^{-16} (ref. 17). Similar precise testings of octave-spanning combs have been performed recently by Scott Diddams and co-workers at the National Institute for Standards and Technology (NIST) in Boulder, Colorado⁴⁸.

The all-optical clock

An optical clock consists, like any other clock, of an oscillator that defines the ticks in time and a counter that keeps track of these cycles. In a caesium clock, for example, the oscillations are those of the electron precessing around the spin of the nucleus. An electronic counter advances the second hand every time the counter has completed 9,192,631,770 oscillations. This number was chosen when the SI second was redefined for the last time in 1967. If we consider the history of timekeeping and compare clocks as different as sun dials, pendulum clocks and quartz clocks it is obvious that they get more accurate as the oscillation frequency increases³. This is simply because a higher oscillation frequency divides time into smaller pieces.

The caesium 9.2-GHz oscillator has been used since the end of the 1950s. But operating at much higher frequencies has now become possible after tremendous advances in laser spectroscopy in the 1970s (refs 1,2) that resulted in trapped ions⁴⁹ and trapped atoms⁵⁰ standards in the 1980s. Systematic uncertainties can be reduced to 10^{-18} for some of these standards⁴¹. When it became possible to count these optical oscillations with harmonic frequency chains in the late 1960s (ref. 4), physicists started to think seriously of running an optical clock. However, working with these counters was so tedious that most of the chains never reached the stage where they could operate continuously, even for minutes. So these chains were used only to calibrate some chosen frequencies, such as iodine- or methane-stabilized He-Ne lasers, which could then be reproduced in other laboratories that could not afford the huge efforts and resources of setting up a harmonic frequency chain. The calibrated lasers were used mostly as wavelength references in interferometers for the realization of the metre, and in some scientific experiments as frequency references.

With the development of the femtosecond frequency synthesizer, a reliable, running optical clock has now come into reach. Owing to its simplicity it can already run for hours and may eventually become a genuine turn-key system.

Currently, the set-up that probably comes closest to an 'optical clock' is operated at NIST³ and uses a transition at 1,064 THz in a trapped single mercury ion. The measurement of the stability of this clock is limited essentially by comparison with the NIST hydrogen maser ensemble, which is one of the most stable radio frequencies available. Other national standard institutions are making progress in the same direction, including work on Yb⁺ by Physikalisch-Technische Bundesanstalt²⁶ and on Sr⁺ by the National Research Council of Canada²⁵ and the National Physics Laboratory (Teddington,

UK)²⁴. At MPQ, Herbert Walther's research group are preparing a clock based on a trapped indium ion⁴¹.

We believe that the development of accurate optical frequency synthesis marks only the beginning of an exciting new period of ultra-precise physics, and that optical clocks will open a new window to nature where we can expect new discoveries and phenomena. One example is the quest for natural constants that vary spatially or that would drift slowly in time as the Universe evolves, as discussed by some theoreticians^{51,52}. Until now, experimental laboratory tests have not been able to detect such behaviour⁵³, but recent experiments are believed to provide evidence for a cosmological evolution of the fine-structure constant⁵⁴. Additionally, these clocks may help to refine general relativity, which still poses one of the main problems in physics, as it refuses proper quantization. The precision of the best test is 'only' as good as 7 parts in 10^5 (ref. 55), which might be the reason why small corrections attributable to a quantized theory have not yet been discovered. Finally, industrial applications such as satellite navigation, communication and network synchronization could benefit greatly from this technology. □

- Bloembergen, N. (ed.) *Non-linear Spectroscopy* (Proc. Int. School Phys. "Enrico Fermi") (North Holland, Amsterdam, 1977).
- Hänsch, T. W. & Inguscio, M. (eds) *Frontiers in Laser Spectroscopy* (Proc. Int. School Phys. "Enrico Fermi") (North Holland, Amsterdam, 1994).
- Diddams, S. A. *et al.* An optical clock based on a single trapped ¹⁹⁹Hg ion. *Science* **293**, 825–828 (2001).
- Evenson, K. M., Wells, J. S., Petersen, F. R., Danielson, B. L. & Day, G. W. Accurate frequencies of molecular transitions used in laser stabilization: the 3.39- μ m transition in CH₄ and the 9.33- and 10.18- μ m transitions in CO₂. *Appl. Phys. Lett.* **22**, 192–195 (1973).
- Schnatz, H., Lipphardt, B., Helmcke, J., Riehle, F. & Zinner, G. First phase-coherent frequency measurement of visible radiation. *Phys. Rev. Lett.* **76**, 18–21 (1996).
- Udem, Th. *et al.* Phase-coherent measurement of the hydrogen 1S–2S transition frequency with an optical frequency interval divider chain. *Phys. Rev. Lett.* **79**, 2646–2649 (1997).
- Schwob, C. *et al.* Optical frequency measurement of the 2S–12D transitions in hydrogen and deuterium: Rydberg constant and Lamb shift determinations. *Phys. Rev. Lett.* **82**, 4960–4963 (1999); erratum *Phys. Rev. Lett.* **86**, 4193 (2001).
- Bernard, J. E. *et al.* Cs-based frequency measurement of a single trapped ion transition in the visible region of the spectrum. *Phys. Rev. Lett.* **82**, 3228–3231 (1999).
- Udem, Th. *Phasenkohärente optische Frequenzmessungen am Wasserstoffatom*. Thesis, Ludwig-Maximilians Univ. (1997).
- Reichert, J., Holzwarth, R., Udem, Th. & Hänsch, T. W. Measuring the frequency of light with mode-locked lasers. *Opt. Commun.* **172**, 59–68 (1999).
- Udem, Th., Reichert, J., Holzwarth, R. & Hänsch, T. W. Accurate measurement of large optical frequency differences with a mode-locked laser. *Opt. Lett.* **24**, 881–883 (1999).
- Udem, Th., Reichert, J., Holzwarth, R. & Hänsch, T. W. Absolute optical frequency measurement of the cesium D1 line with a mode-locked laser. *Phys. Rev. Lett.* **82**, 3568–3571 (1999).
- Jones, D. J. *et al.* Carrier-envelope phase control of femtosecond mode-locked lasers and direct optical frequency synthesis. *Science* **288**, 635–639 (2000).
- Diddams, S. A. *et al.* Direct link between microwave and optical frequencies with a 300 THz femtosecond laser comb. *Phys. Rev. Lett.* **84**, 5102–5105 (2000).
- Reichert, J. *et al.* Phase coherent vacuum-ultraviolet to radio frequency comparison with a mode-locked laser. *Phys. Rev. Lett.* **84**, 3232–3235 (2000).
- Niering, M. *et al.* Measurement of the hydrogen 1S–2S transition frequency by phase coherent comparison with a microwave cesium fountain clock. *Phys. Rev. Lett.* **84**, 5496–5499 (2000).
- Holzwarth, R. *et al.* Optical frequency synthesizer for precision spectroscopy. *Phys. Rev. Lett.* **85**, 2264–2267 (2000).
- Udem, Th. *et al.* Absolute frequency measurements of the Hg⁺ and Ca optical clock transitions with a femtosecond laser. *Phys. Rev. Lett.* **86**, 4996–4999 (2001).
- Stenger, J. *et al.* Phase-coherent frequency measurement of the Ca intercombination line at 657 nm with a Kerr-lens mode-locked femtosecond laser. *Phys. Rev. A* **63**, 021802–1–021802–4 (2001).
- Pokasov, P. V. *et al.* in *Proc. Sixth Symp. Freq. Standards Metrol.* (ed. Gill, P.) 510–512 (World Scientific, Singapore, 2002).
- Nevsky, A. Yu. *et al.* Frequency comparison and absolute frequency measurement of I₂ stabilized lasers at 532 nm. *Opt. Comm.* **263**, 192–272 (2001).
- Holzwarth, R. *et al.* Absolute frequency measurement of iodine lines with a femtosecond optical synthesizer. *Appl. Phys. B* **73**, 269 (2001).
- Ye, J. *et al.* Accuracy comparison of absolute optical frequency measurement between harmonic-generation synthesis and a frequency division femtosecond-comb. *Phys. Rev. Lett.* **85**, 3797–3800 (2000).
- Lea, S. N. *et al.* in *Proc. Sixth Symp. Freq. Standards Metrol.* (ed. Gill, P.) 144–151 (World Scientific, Singapore, 2002).
- Dubé, P., Marmet, L., Bernard, J. E., Siemsen, K. J. & Madej, A. A. in *Proc. Sixth Symp. Freq. Standards Metrol.* (ed. Gill, P.) 489–491 (World Scientific, Singapore, 2002).
- Stenger, J., Tamm, Ch., Haverkamp, N., Weyers, S. & Telle, H. R. Absolute frequency measurement of the 435.5 nm ¹⁷¹Yb⁺-clock transition with a Kerr-lens mode-locked femtosecond laser. *Opt. Lett.* **26**, 1589–1591 (2001).
- von Zanthier, J. *et al.* Absolute frequency measurement of the In⁺ clock transition with a mode-locked laser. *Opt. Lett.* **25**, 1729–1731 (2000).
- Apolonski, A. *et al.* Controlling the phase evolution of few-cycle light pulses. *Phys. Rev. Lett.* **85**, 740–743 (2000).

29. Telle, H. R., Steinmeyer, G., Dunlop, A. E., Sutter, D. H. & Keller, U. Carrier-envelope offset phase control: a novel concept for absolute optical frequency measurement and ultrashort pulse generation. *Appl. Phys. B* **69**, 327–332 (1999).
30. Xu, L. *et al.* Route to phase control of ultrashort light pulses. *Opt. Lett.* **21**, 2008–2010 (1996).
31. Paulus, G. G. *et al.* Evidence of 'absolute-phase' phenomena in photoionization with few-cycle laser pulses. *Nature* **414**, 182–184 (2001).
32. Drescher M. *et al.* X-ray pulses approaching the attosecond frontier. *Science* **291**, 1923–1927 (2001).
33. Eckstein, J. N., Ferguson, A. I. & Hänsch, T. W. High-resolution two-photon spectroscopy with picosecond light. *Phys. Rev. Lett.* **40**, 847–850 (1978).
34. Chebotayev, V. P. & Ulybin, V. A. Synchronization of atomic quantum transitions by light pulses. *Appl. Phys.* **50**, 1–5 (1990).
35. Kane, D. M., Bramwell, S. R. & Ferguson, A. I. FM dye lasers. *Appl. Phys. B* **39**, 171–178 (1986).
36. Telle, H. R. in *Frequency Control of Semiconductor Lasers* (ed. Ohtsu, M.) 137–167 (Wiley, New York, 1996).
37. Hänsch, T. W. in *The Hydrogen Atom* (eds Bassani, G. F., Inguscio, M. & Hänsch, T. W.) 93–102 (Springer, Berlin, 1989).
38. Telle, H. R., Meschede, D. & Hänsch, T. W. Realization of a new concept for visible frequency division: phase-locking of harmonic and sum frequencies. *Opt. Lett.* **15**, 532–534 (1990).
39. Wicht, A., Hensley, J. M., Sarajlic, E. & Chu, S. A preliminary measurement of $\hbar/M_{\text{In}}$ with atom interferometry, in *Proc. Sixth Symp. Freq. Standards Metrol.* (ed. Gill, P.) (World Scientific, Singapore, in the press).
40. Hensley, J. M. *A Precision Measurement of the Fine Structure Constant*. Thesis, Stanford Univ. (2001).
41. Becker, Th., von Zanthier, J. & Nevsky, A. Yu. High-resolution spectroscopy of a single In^+ ion: progress towards an optical frequency standard. *Phys. Rev. A* **63**, 051802-1–051802-4 (2001).
42. Udem, Th., Holzwarth, R. & Hänsch, T. W. in *Proceeding of Joint Meeting of the 13th European Frequency and Time Forum and 1999 IEEE International Frequency Control Symposium, Besancon, France, 13-16 April 1999* 620–625 (IEEE Publications, 1999).
43. Knight, J. C., Birks, T. A., Russell, P. St. J. & Atkin, D. M. Endlessly single-mode photonic crystal fibre. *Opt. Lett.* **22**, 961–964 (1996).
44. Wadsworth, W. J. *et al.* Soliton effects in photonic crystal fibres at 850 nm. *Electron. Lett.* **36**, 53 (2000).
45. Ranka, J. K., Windeler, R. S. & Stentz, A. J. Visible continuum generation in air-silica microstructure optical fibres with anomalous dispersion at 800 nm. *Opt. Lett.* **25**, 25–28 (2000).
46. Birks, T. A., Wadsworth, W. J. & Russell, P. St. J. Supercontinuum generation in tapered fibres. *Opt. Lett.* **25**, 1415–1417 (2000).
47. Ell, R. *et al.* Generation of 5-fs pulses and octave-spanning spectra directly from a Ti:sapphire laser. *Opt. Lett.* **26**, 373–375 (2001).
48. Diddams, S. A., Hollberg, L., Ma, L. S. & Robertson, L. A femtosecond-laser-based optical clockwork with instability 6.3×10^{-16} in 1 s. *Opt. Lett.* **27**, 58 (2002).
49. Madej, A. A. & Bernard, J. E. in *Frequency Measurement and Control* (ed. Luiten, A. N.) 153–194 (Springer, Berlin, 2001).
50. Riehle, F. & Helmcke, J. in *Frequency Measurement and Control* (ed. Luiten, A. N.) 95–129 (Springer, Berlin, 2001).
51. Dirac, P. A. M. The cosmological constants. *Nature* **139**, 323 (1937).
52. Karshenboim, S. G. Some possibilities for laboratory searches for variations of fundamental constants. *Can. J. Phys.* **78**, 639–678 (2000).
53. Salomon, Ch. *et al.* in *Atomic Physics 17: XVII Int. Conf. Atom. Phys.; ICAP 2000* (eds Arimondo, E., De Natale, P. & Inguscio, M.) 23–40 (AIP Conf. Proc. Vol. 551) (American Institute of Physics, 2001).
54. Webb, J. K. *et al.* Further evidence for cosmological evolution of the fine structure constant. *Phys. Rev. Lett.* **87**, 091301-1–091301-4 (2001).
55. Vessot, R. F. C. *et al.* Test of relativistic gravitation with a space-borne hydrogen maser. *Phys. Rev. Lett.* **45**, 2081–2084 (1980).
56. Ferguson, A. I., Eckstein, J. N. & Hänsch, T. W. Polarization spectroscopy with ultrashort light pulses. *Appl. Phys.* **18**, 257 (1979).
57. Wineland, D. J., Bergquist, J. C., Itano, W. M., Diedrich, F. & Weimer, C. S. in *The Hydrogen Atom* (eds Bassani, G. F., Inguscio, M. & Hänsch, T. W.) 123–133 (Springer, Berlin, 1989).

Acknowledgements

We thank our collaborators P. Lemonde, G. Santarelli, M. Abgrall, P. Laurent and A. Clairon (BNM-LPTF, Paris), C. Salomon (ENS, Paris), J. Knight, W. Wadsworth, T. Birks and P. St J. Russell (University of Bath), and J. Hall, S. Diddams, J. Ye and L. Hollberg (NIST, Boulder, Colorado) for the excellent teamwork and stimulating discussions.

Quantum information processing with atoms and photons

C. Monroe

FOCUS Center and Department of Physics, University of Michigan, Ann Arbor, Michigan 48109-1120, USA (e-mail: crmonroe@umich.edu)

Quantum information processors exploit the quantum features of superposition and entanglement for applications not possible in classical devices, offering the potential for significant improvements in the communication and processing of information. Experimental realization of large-scale quantum information processors remains a long-term vision, as the required nearly pure quantum behaviour is observed only in exotic hardware such as individual laser-cooled atoms and isolated photons. But recent theoretical and experimental advances suggest that cold atoms and individual photons may lead the way towards bigger and better quantum information processors, effectively building mesoscopic versions of ‘Schrödinger’s cat’ from the bottom up.

In 1948, Claude Shannon discovered how to quantify information¹ — a result so fundamental and revolutionary that in hindsight it is surprising it had not been formulated earlier. Shannon’s bit, or binary digit, became the fundamental unit of information, providing a metric for comparing forms of information and optimizing the amount of resources needed to faithfully convey a given amount of information, even in the presence of noise.

Shannon’s pioneering work not uncoincidentally prefaced the information age. Bits were found in nature, from unwieldy vacuum tubes in the 1940s to the modern semiconductor transistors of less than 10^{-5} cm in size. Under this impressive progression of technology, we have enjoyed an exponential growth in computing power and information processing speed given by the familiar ‘Moore’s Law’, where computer chips have doubled in density every year or two. Unfortunately, the days (or years) of Moore’s Law are numbered. As bits continually shrink in size, they will eventually approach the size of individual molecules — by the year 2020 if the current growth continues. At these nanometre-length scales, the laws of quantum mechanics begin to hold sway. In fact, classical bits can still be stored and manipulated here, as verified by the ground-breaking theoretical work of Paul Benioff^{2,3} and Richard Feynman⁴ in the early 1980s, and foreseen in Feynman’s remarkable 1959 charge towards nanotechnology, *There’s Plenty of Room at the Bottom*⁵. But even if we could somehow reach the bottom, where quantum states of every atom in a computer chip host classical bits, there would be no more room for further gains without splitting the atom.

Quantum rules

If we instead focus on the pure quantum features of a collection of atom-sized bits, we indeed find more room. For an isolated quantum system, the fundamental unit of information is the quantum bit or ‘qubit’. Qubits are just quantum two-level systems such as the spin of an electron or the polarization of a photon, and can be prepared in a coherent superposition state of 0 and 1:

$$|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle \quad (1)$$

Here, α and β are the complex amplitudes of qubit states $|0\rangle$ and $|1\rangle$, and the superposition is resolved into either

definite state $|0\rangle$ or $|1\rangle$ upon measurement with respective probabilities $|\alpha|^2$ and $|\beta|^2 = 1 - |\alpha|^2$. In one sense, this is just a single bit of information. But in another sense, the continuous amplitudes α and β carry an infinite amount of information, similar to analogue information carriers such as the continuous voltage stored on capacitors. However, analogue systems are known to suffer from the cumulative build-up of noise, as opposed to digital standards like transistor–transistor logic that latch to their high or low levels through constant measurement and feedback. Quantum bits are similarly vulnerable to analogue noise⁶, but they can offer much more in return: quantum entanglement. In a classical analogue system, we need N capacitors to store N continuous voltages. But with N qubits, the most general state is specified by 2^N independent amplitudes $\gamma_0, \dots, \gamma_{2^N-1}$

$$|\Psi\rangle = \gamma_0|0_1 0_2 0_3 \dots 0_N\rangle + \gamma_1|0_1 0_2 0_3 \dots 1_N\rangle + \dots + \gamma_{2^N-1}|1_1 1_2 1_3 \dots 1_N\rangle \quad (2)$$

A collection of qubits therefore has the potential for storing exponentially more information than a comparable collection of classical information carriers. (Of course, the trick is extracting this information following a measurement, as discussed in the below applications.) The above state is entangled, as it is not generally separable into a product of individual qubit states. The implicit interconnects from entanglement give a quantum information processor its power (ref. 7; and see Box 1).

Atoms and photons: good quantum hardware

The chief hardware requirements for a quantum information processor are⁸:

1. The quantum system (that is, a collection of qubits) must be initialized in a well-defined state such as $|0_1 0_2 0_3 \dots 0_N\rangle$.
2. Arbitrary unitary operators must be available and controlled to launch the initial state to an arbitrary entangled state (equation (2)).
3. Measurements of the qubits must be performed with high quantum efficiency.

The first two requirements demand that the qubits are well isolated from the environment to ensure pure initial quantum states and to preserve their superposition character, but they must also interact strongly between one another in order to become entangled. On the other hand,

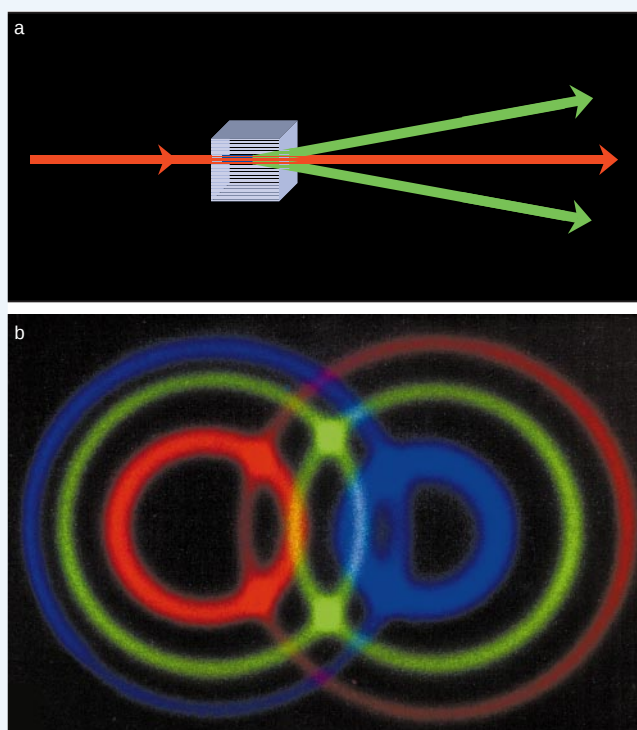


Figure 1 Optical parametric down-conversion. **a**, An ultraviolet photon incident on a nonlinear crystal can sometimes split spontaneously into two daughter photons. These photons are emitted on opposite sides of the pump beam, along two cones, one of which has horizontal polarization, the other of which has vertical polarization. **b**, Along the optical axis, several cone pairs can be seen. Photon pairs emitted along the intersections of the cones are entangled in polarization. (Image courtesy of A. Zeilinger, University of Vienna.)

the last stipulation requires the strongest possible interaction with the environment to be switched on at will.

These stringent hardware requirements unfortunately rule out most known physical systems. Conventional solid-state architectures such as silicon are ideal for classical information for the same reason they are unsuitable for quantum information; as discussed above, their stability is given by the latching of logic levels, or continuous monitoring by the environment⁶. The most attractive candidates for quantum information processors currently come from the area of atomic physics and quantum optics⁹. Here, individual atoms and photons are manipulated in a controlled environment with well-understood couplings, offering environmental isolation that is unsurpassed in other physical systems.

Complex entangled states such as equation (2) can be generated by sequentially applying quantum logic-gate primitives to smaller numbers of qubits, similar to the use of sequences of universal logic gates such as NAND ('not-and') in classical computation. In fact, it has been shown that arbitrary entangled states can be generated from sequences of simple qubit logic gates acting on any one or two qubits at a time^{10,11}. This sequential model of quantum computing fits naturally with quantum optical and atomic systems, where interactions are generally weak and typically involve at most two qubits.

Quantum information processing requires qubits to behave as quantum memories for long-term storage and for many applications to behave as quantum transmitters for long-distance communication. Cold and localized individual atoms are the natural choice for qubit memories and sources of local entanglement for quantum information processing. The stability of quantum states in cold atoms is unrivalled, evidenced by the fact that such quantum systems currently host the world's best frequency standards (see review in this issue by Udem, Holzwarth and Hänsch, pages 233–237). Individual

Box 1

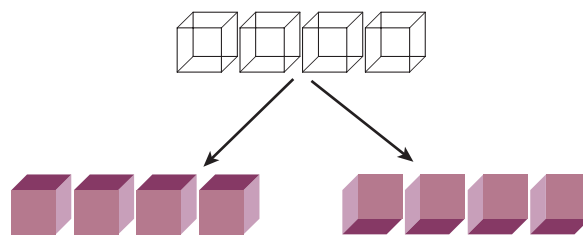
Entanglement, Schrödinger's cat and decoherence

Multiple quantum bits can be prepared in entangled states that are not able to be factored into a product of individual qubit states. An interesting class of entangled states is the 'Schrödinger-cat' states of N quantum bits, written as

$$|\Psi\rangle_{\text{cat}} = |0_1 0_2 0_3 \dots 0_N\rangle + |1_1 1_2 1_3 \dots 1_N\rangle \approx |\text{'live cat'}\rangle + |\text{'dead cat'}\rangle$$

This state is known as a mesoscopic Schrödinger-cat state, as the constituents of the superposition are as far apart as possible, analogous to the live and dead feline. As Schrödinger himself noted, describing a macroscopic object such as a cat with quantum terms is of course ridiculous, but for mesoscopic systems where N is not too large, the above state is useful in grasping large-scale entanglement.

The qubits in a Schrödinger-cat state are perfectly correlated, yet each qubit is entirely random when considered alone. This hidden wiring of entanglement, responsible for gains in quantum information processing, can be roughly visualized in the following analogy⁸⁹. Consider $N = 4$ adjacent three-dimensional cubes drawn with ambiguous perspective.



After a brief moment, your mind typically locks onto one of the two perspectives, and you will almost certainly see all cubes emerge with the same perspective. This is roughly analogous to a measurement of an entangled superposition.

The Schrödinger-cat state also highlights the difficulty in maintaining complex entangled quantum superpositions. If just one of the N qubits gets measured by the environment, every qubit loses its coherence. Decoherence makes it more difficult to engineer large entangled states, with the coherence survival probability decaying exponentially with the number of qubits⁹⁰. Extrapolating to $N \sim 10^{28}$ qubits in a real Schrödinger's cat, we find that the superposition of live and dead would almost instantaneously collapse into one or the other.

photons, on the other hand, are the natural source for the communication of quantum information, as they can traverse large distances through the atmosphere or optical fibres with minimal disturbance.

Quantum communication

Imagine two distant parties, Alice and Bob, share two entangled qubits analogous to an Einstein–Podolsky–Rosen (EPR) state¹²

$$|\Psi_{AB}\rangle = |0_A 0_B\rangle + |1_A 1_B\rangle \quad (3)$$

What can Alice and Bob do with their entangled pair? At first glance, they might try to exploit the strong correlations in their entangled states for direct (and superluminal) communication. A measurement by Alice or Bob immediately conveys the state of the other's qubit, but such measurements would yield only a sequence of perfectly random (yet correlated) bits, which carry no information according to Shannon. On the other hand, correlated random strings

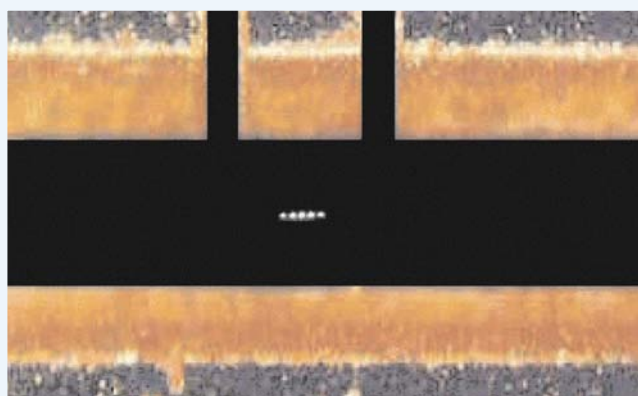


Figure 2 A crystal of five atomic beryllium ions (small white dots at centre) confined in a radio-frequency ion trap. The ions balance their mutual Coulomb repulsion with the confining force of electric fields generated from the surrounding electrodes (brown). The ions strongly fluoresce under the application of appropriate laser radiation near 313 nm. The horizontal electrode gap is about 0.2 mm and the ion-ion spacing is $\sim 5 \mu\text{m}$. (Image courtesy of NIST, Boulder.)

are extremely useful for cryptographic key distribution, as in the classic 'one-time pad'¹³.

Quantum cryptographic key distribution

Charles Bennett and Gilles Brassard developed the first protocols for quantum cryptographic key distribution (QKD) in 1984 (ref. 14), following earlier ideas of Steven Wiesner¹⁵. This scheme used the transmission of non-orthogonal states of single-photon qubits¹⁶, with security derived from the impossibility of an eavesdropper distinguishing the two states without being detected (on average). Several groups have implemented these protocols over tens of kilometres in underground fibres^{17–20}. Instead of idealized single-photon sources, these experiments used highly attenuated laser sources, which have a finite probability of emitting multiple photons per pulse. This gives rise to a trade-off between bit communication rate and security, resulting from the potential for eavesdropping on the surplus photons.

The use of entangled qubits can eliminate this trade-off and dramatically improve QKD. For example, by encoding qubits in orthogonal states of photon polarization in the form of equation (3), the measured presence of one photon at Alice indicates that Bob has

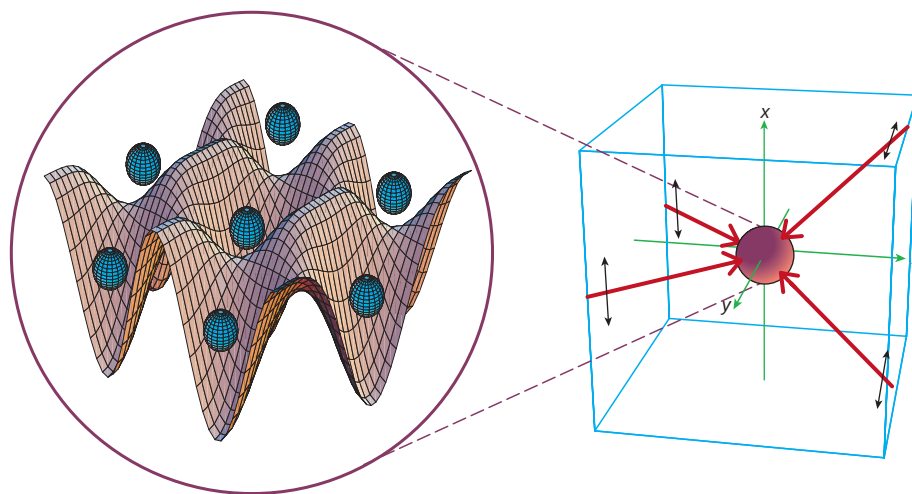
exactly one photon, which can be used to improve the security of the above single-qubit QKD scheme. Entangled qubits also offer an alternative QKD protocol, first pointed out by Artur Ekert in 1991 (ref. 21), and implemented recently by groups at Vienna University²², Los Alamos National Laboratory²³ and the University of Geneva²⁴. Here, the security is derived from the fact that a potential eavesdropper disturbs the entanglement shared by Alice and Bob.

These experiments used nonlinear optical parametric down-conversion (PDC) as their source of EPR-like entangled pairs. Ultraviolet pump photons are directed into a nonlinear crystal that gives rise to a small probability of down-conversion into pairs of visible or infrared daughter photons, as depicted in Fig. 1. These daughter photons always appear simultaneously, and can be entangled in their polarization degrees of freedom with appropriate phase-matching and mode selection. Unfortunately, PDC photon sources have several drawbacks in quantum information applications. Similar to the case of a highly attenuated laser source, there is a small probability of emitting pairs of down-converted photons that again offers an avenue for potential eavesdropping. Furthermore, PDC suffers from poor efficiency, as the probability of down-conversion per pump photon is typically below 10^{-10} (ref. 25). The presence of an entangled pair can be inferred only after a measurement (post-selection), so such states cannot be cascaded for subsequent processing to produce large-scale entangled photon states. Nevertheless, PDC sources are very useful for demonstrating rudimentary quantum communication protocols such as QKD.

Quantum teleportation

How can qubits be transmitted between remote locations? One method is to simply move the qubits through space, similar to the transmission of classical bits with electrical currents. However, this can be very difficult in practice, especially over long distances, as it must be done without disturbing or measuring the qubit state. Quantum teleportation, discovered by Bennett in 1993 (ref. 26), offers an alternative method of transmission without physical contact²⁷, as described in Box 2. The drawback is that the sender and receiver must already possess entangled qubits, a situation that may be as difficult as physically moving qubits from A to B in the first place, as entanglement always originates between nearby qubits. The real power of quantum teleportation is that it allows quantum information to be conveyed over a channel that may be unsuitable for direct physical transmission at the time of communication, or a channel that cannot reliably host the form of qubit stored at each end. For example, teleportation might allow quantum states in

Figure 3 Schematic of an optical lattice. A pattern of crossed laser beams (right) creates an array of potential wells that can confine individual laser-cooled cold atoms (left). The atom-atom separation is of the order of the optical wavelength. (Image courtesy of S. Rolston, NIST.)



atoms to be communicated between remote locations that possess previously entangled photons.

Groups from the universities of Innsbruck²⁸ and Rome²⁹ were the first to demonstrate certain critical aspects of quantum teleportation in 1997. In these experiments, polarization qubits of single photons were teleported between spatially separated locations using PDC. The low efficiency of these sources precluded the teleportation of larger numbers of qubits or the recycling of qubits^{30,31}. Shortly thereafter, a team at the California Institute of Technology (CalTech) efficiently teleported the continuous quantum state of a single-mode optical field³², relying on the deterministic entanglement of a pair of field modes, demonstrating a net fidelity of 58% without the need for post-selection.

Will we ever be able to teleport objects with more than just a single quantum degree of freedom? The answer is yes, as long as the three quantum hardware requirements are satisfied. But what about teleporting truly macroscopic objects with perhaps 10^{28} effective qubits? This prospect, about as likely as preparing a real Schrödinger's cat both alive and dead, is clearly "ridiculous" as Schrödinger himself said³³. Instead, it is more fruitful to consider what we can do with a more modest system of perhaps 300 qubits, where the $\sim 10^{90}$ available quantum states are still more than the number of particles in the Universe.

Quantum computing

In 1985, David Deutsch showed how quantum superposition and entanglement could be harnessed to process information more efficiently than any classical machine³⁴. Deutsch's 'quantum parallelism' allows an N -qubit quantum computer to operate on quantum superposition states of all 2^N inputs. Extraction of information in a quantum computer is nontrivial, because simply measuring a quantum state such as equation (2) with an exponential number of nonzero amplitudes will yield a highly random result. In certain algorithms, however, appropriate quantum logic gates cause the amplitudes to interfere so that only a few amplitudes survive in the end. Following a measurement (or a limited number of repeated measurements on identical runs), the result (or distribution of results) can depend on a global property of all 2^N inputs.

The best-known example of this procedure is Peter Shor's 1994 factoring algorithm³⁵, where a quantum computer would be able to factor large numbers exponentially faster than any known classical computer algorithm. Fast quantum number factoring has profound implications in cryptanalysis, as many popular encryption algorithms such as RSA (Rivest–Shamir–Adleman)³⁶ rely on the inability to factor large numbers. In 1996, Lov Grover discovered a quantum algorithm that would search an unsorted database of M elements in a time that scales roughly $M^{1/2}$ faster than any possible classical search algorithm³⁷. These discoveries have ushered a flurry of theoretical activity searching for further useful quantum computer applications.

A key development in 1995 was quantum error correction^{38,39}, an extension of Shannon's famous 1948 noisy coding theorem¹ to the quantum realm. Quantum error correction detects and corrects slight quantum gate errors or weak interactions with the environment through redundant encoding of qubits. This means that experiments need not be perfect; they require only that noise levels be below a certain threshold, with a trade-off between the amount of extra qubits required and the error threshold level. According to some models⁴⁰, arbitrary-length quantum computation can proceed error-free with only a polynomial overhead in time and space, so long as the error probability per fundamental operation is kept below about 10^{-5} . What follows is a review of the most promising quantum computer architectures that may someday reach this level of fidelity.

Ion traps

Laser-cooled and trapped atomic ions represent one of the most attractive candidates for a large-scale quantum computer^{41–44}. Here, electromagnetic fields confine individual atoms in free space in a

Box 2

Quantum teleportation

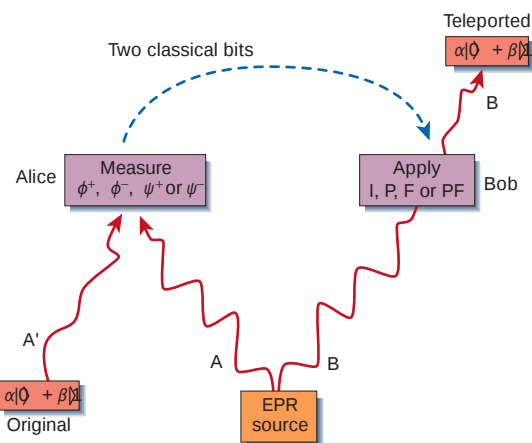
Quantum teleportation, the 'disembodied transport' of a quantum state from one site to another⁹¹, is conceptually simple, as outlined in the figure below. Imagine that Alice and Bob each already possess one of two qubits prepared in an entangled state such as equation (3). If Alice prepares another qubit A' in the arbitrary state $\alpha|0\rangle_{A'} + \beta|1\rangle_{A'}$, the overall state of the three qubits is

$$\begin{aligned} |\Psi\rangle &= (\alpha|0\rangle_{A'} + \beta|1\rangle_{A'}) (|0\rangle_A |0\rangle_B + |1\rangle_A |1\rangle_B) \\ &= |\phi^+\rangle_{A'A} (\alpha|0\rangle_B + \beta|1\rangle_B) + |\phi^-\rangle_{A'A} (\alpha|0\rangle_B - \beta|1\rangle_B) \\ &\quad + |\psi^+\rangle_{A'A} (\alpha|1\rangle_B + \beta|0\rangle_B) + |\psi^-\rangle_{A'A} (\alpha|1\rangle_B - \beta|0\rangle_B) \end{aligned} \quad (5)$$

where the four 'Bell' states of the Alice's qubits are defined as

$$\begin{aligned} |\phi^\pm\rangle &\equiv |0_A 0_{A'}\rangle \pm |1_A 1_{A'}\rangle \\ |\psi^\pm\rangle &\equiv |0_A 1_{A'}\rangle \pm |1_A 0_{A'}\rangle \end{aligned} \quad (6)$$

To pass the quantum information of Alice's qubit (the amplitudes α and β) to Bob, Alice makes a perfect projection measurement of her qubit pair onto this complete basis of Bell states. This can be done with simple operations of quantum logic gates on qubits A and A' along with an efficient measurement of both qubits, as depicted in the figure. This measurement immediately projects Bob's lone qubit onto one of the four states correlated with the state measured by Alice given in equation (5), resulting in his qubit now carrying the quantum information α and β . After Alice communicates which Bell state she measured to Bob using classical means (for example, a phone call), Bob performs a prescribed local qubit manipulation on his one qubit to replicate the initial state $\alpha|0\rangle + \beta|1\rangle$ in his qubit. For example, if Alice reports a measurement of $|\phi^+\rangle$, then Bob does nothing to his qubit (operation I). If Alice instead reports a measurement of $|\psi^-\rangle$, then Bob flips his qubit (operation F) and additionally adds a π -phase shift (operation P). In this way, the qubit $\alpha|0\rangle + \beta|1\rangle$ is 'teleported' from Alice to Bob without disturbing or measuring the qubit.



vacuum chamber, and when multiple ions are confined and laser-cooled, they form simple stationary crystal structures given by the balance of the external confining force of the trap with the mutual repulsion of the atoms (Fig. 2). Qubits are stored in internal electronic states of the atoms, typically the same long-lived hyperfine states that are used in atomic clocks. When appropriate laser radiation is directed to the atomic ions, qubit states can be mapped coherently onto the quantum state of collective motion of the atoms and subsequently mapped to other atoms. A single normal mode of collective

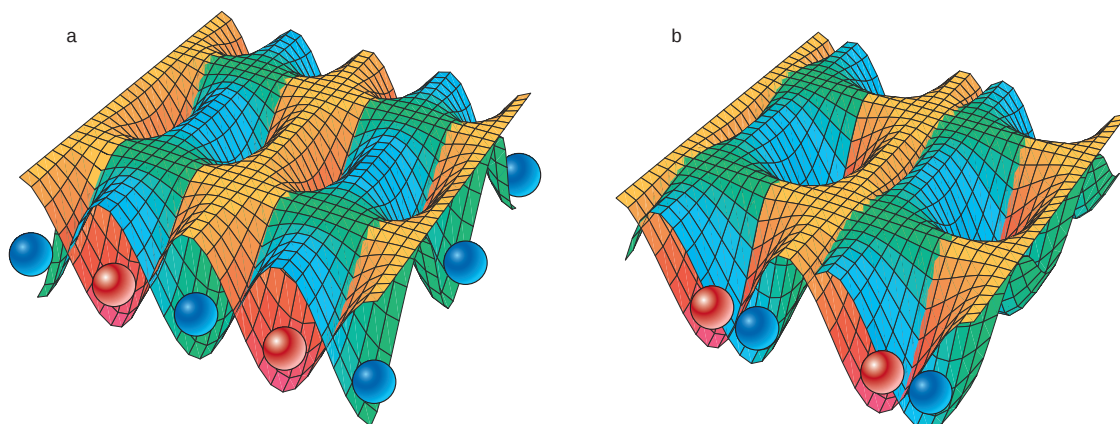


Figure 4 Optical lattice potentials. **a**, Optical lattice potentials can depend on the internal state (red or blue) of the confined atom qubits. **b**, These complementary lattice potentials can be controlled by varying the polarization of one of the lattice light fields,

shifting the interference pattern and bringing atoms together in pairs for quantum logic gates. (Image courtesy of I. Deutsch, University of New Mexico.)

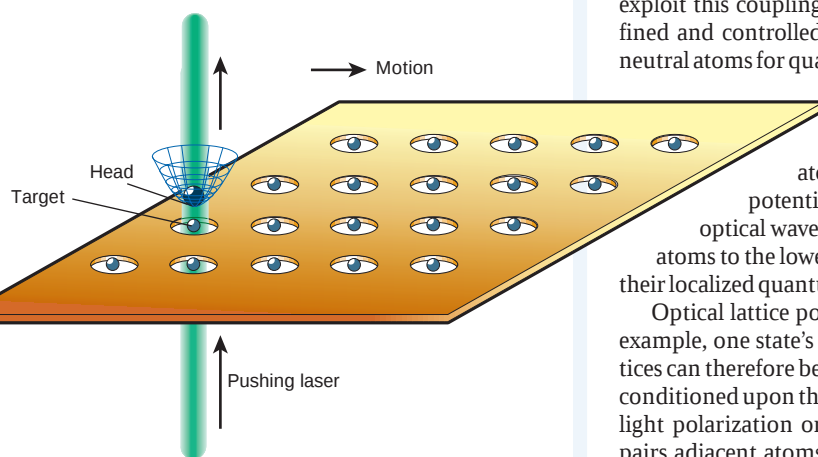
crystal motion thus behaves as a quantum data-bus, allowing quantum information to be shared and entangled between remote atomic qubits in the crystal. Finally, the internal states of individual trapped ions can be measured with nearly 100% quantum efficiency⁴⁵ by applying appropriate laser radiation and collecting fluorescence — one qubit state is bright and the other is dark.

Quantum logic gates have been demonstrated with up to four trapped atomic ion qubits at the National Institute of Standards and Technology (NIST) at Boulder, Colorado^{46,47}. Although this scheme is in principle scalable to arbitrarily large numbers of qubits, the main problems deal with the collective motion of the atoms. The quantum gate speed is limited to the frequency of motion, typically in the sub-megahertz range. As more qubits are added to the collection, the density of motional states balloons, and isolation of a single mode of motion (for example, the centre-of-mass) becomes even more slow and difficult. Moreover, external noisy electric fields tend

to compromise the motional coherence of large numbers of trapped atomic ions⁴⁸.

A promising approach that attacks both problems is the quantum charged-coupled device (CCD), where individual atomic ions are entangled as above, but only among a small collection (under 10) of atomic ions in an ‘accumulator’⁴³. Scaling to larger numbers is accomplished by physically swapping individual atoms between the accumulator and a ‘memory’ reservoir of trapped atom qubits. This can be done quickly with externally applied electric fields in elaborate ion-trap electrode geometries. The key features of the quantum CCD are that the ion shuttling can be done without perturbing the internal qubits, and the motional quantum state of the ions factors from the internal qubit states following quantum gate operation. In order to quench the motional energy from rapid shuttling and allow subsequent logic gates, ancillary ions in the accumulator can be laser-cooled between gate operations. The qubit ions are thereby sympathetically cooled⁴⁹ through their strong Coulomb interaction with these extra refrigerator ions. Sympathetic cooling can also eliminate motional decoherence during logic gate operation^{43,50,51}.

Figure 5 Scheme for a scalable ion trap and optical-lattice quantum computer. An array of independent (memory) ion traps is coupled through interaction with a roving (head) ion trap. A strong optical-lattice laser field entangles the head ion with a given memory or target ion. (Image courtesy of P. Zoller, University of Innsbruck.)



Cold atoms and dipoles in optical lattices

Neutral atom qubits enjoy a weak coupling to the environment, at the expense of a weak dipole–dipole coupling between each other. To exploit this coupling for entanglement, atoms must be tightly confined and controlled to sub-micron dimensions. A natural host of neutral atoms for quantum information purposes is the optical lattice — an array of cold atoms confined in free space by a pattern of crossed laser beams⁵².

The dipole force between the polarizable atoms and the field results in a regular pattern of potential wells, the spacing of which is of the order of an optical wavelength (Fig. 3). Several groups have laser-cooled atoms to the lowest bound state in lattice wells^{53–55} and controlled their localized quantum wavepacket states^{56,57}.

Optical lattice potentials generally depend upon qubit level (for example, one state’s valley can be another state’s hill). Atoms in lattices can therefore be shifted to nearly overlap with their neighbours, conditioned upon their internal qubit state, by modulating the lattice light polarization or intensity, as shown in Fig. 4. One proposal⁵⁸ pairs adjacent atoms together with one of the atoms prepared in a short-lived excited state. The resultant resonant dipole–dipole interaction, mediated by the shared excitation, allows pair-wise

entanglement. The speed of this quantum gate must be faster than the excited-state decay time, meaning the atoms must be localized and overlapped to well under an optical wavelength. A related proposal would transiently bring adjacent atoms together depending on their internal qubit levels, but instead rely on 'cold collisions' for entangling logic gates⁵⁹. If the atoms are sufficiently cold, then such s-wave collisions will result in a conditional phase shift of their quantum states, or a quantum phase gate.

Another approach exploits the large electric dipole moments of Rydberg atoms⁶⁰ or even simple heteronuclear molecules⁶¹ in an externally applied electric field. For instance, selected atoms held in an optical lattice can be entangled by promoting them to Rydberg states and relying on their mutual dipole–dipole interaction. The interaction strengths over micrometre distances can be in the gigahertz regime, implying very fast quantum gate operation.

One significant drawback to the use of optical lattices in quantum information is that the lattice sites are typically not uniformly packed. However, the same statistical properties of bosons responsible for Bose–Einstein condensation (see review in this issue by Anglin and Ketterle, pages 211–218) can be exploited for better control of atom number in optical lattices. A subtle interplay between a repulsive s-wave scattering interaction between bosonic atoms and the tunnelling of atoms between lattice sites can smooth out the atom-number statistics between wells, a phenomenon known as a Mott-insulator phase transition^{62,63}. Signatures of this behaviour have been observed in an optical lattice loaded with a Bose–Einstein condensate. Researchers at Yale University⁶⁴ measured fluctuations in the number N of atoms per lattice site by coherently combining atoms in adjacent sites in an atom interferometer, and reported a spread in atom number that is a factor of about 40 below the standard $N^{1/2}$ shot-noise level. A group at the Max Planck Institute for Quantum Optics (MPQ) in Garching, near Munich, has recently observed signatures of the small N limit to such 'number-squeezed' states, where exactly one atom resides in each site⁶⁵. This sets the stage for future progress in quantum gate operations using optical lattices.

A further proposal combines the features of optical lattices and ion traps, where individual ions are entangled through a common interaction with a pulsed, high-strength optical lattice⁶⁶. The lattice jerks two ions such that the internal qubit states of each ion are mapped to displaced spatial positions, smearing out their charge distributions⁶⁷. The interaction between these 'engineered' dipoles then allows entangling quantum logic gates that can be much faster than the ion-trap frequency. Moreover, pure quantum states of motion are not necessary, so long as the ions are confined to much less than the ion–ion separation. Finally, the ions need not be close together or even in the same trap, providing a convenient method for scaling to arbitrary numbers of qubits in an array of separate ion traps (Fig. 5).

Quantum networks

As more complex quantum communication protocols are demonstrated, there will be a pressing need to store quantum information in long-term memories. To construct reliable quantum networks of both quantum transmitters and quantum memories, small quantum computers will be needed at the nodes, acting as routers, repeaters and switches.

Degradation of quantum information will be unavoidable when qubits are sent over large distances or over noisy channels. This will require quantum 'repeaters' to be placed periodically along the channel so that qubit errors can be corrected. These repeaters can take the form of error-correcting quantum computers, or stations that share several imperfect entangled qubit pairs with adjacent sites⁶⁸. Here, quantum computer operations at the nodes 'purify' the entangled states so that states can effectively be teleported over remote distances. Remarkably, some of these protocols require local operation fidelities of only 98% for successful transmission⁶⁹.

A key component of such a quantum network is the faithful and coherent conversion of quantum information between different

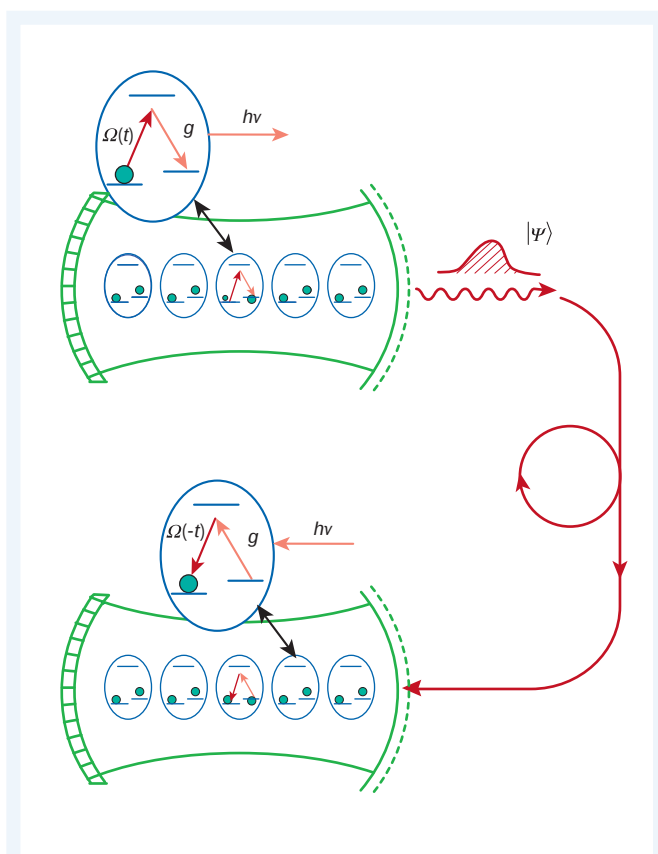


Figure 6 Atom–photon quantum network. A selected atom inside the top optical cavity coherently transfers its internal qubit onto a single-photon qubit in the cavity through the application of a classical laser pulse represented by the coupling $\Omega(t)$. The coupling g is between the single-photon field in the cavity and the atom. The single photon leaks out of the top cavity, only to be caught in the lower cavity by a time-reversed and synchronized classical laser pulse $\Omega(-t)$. (Image courtesy of H. J. Kimble, CalTech.)

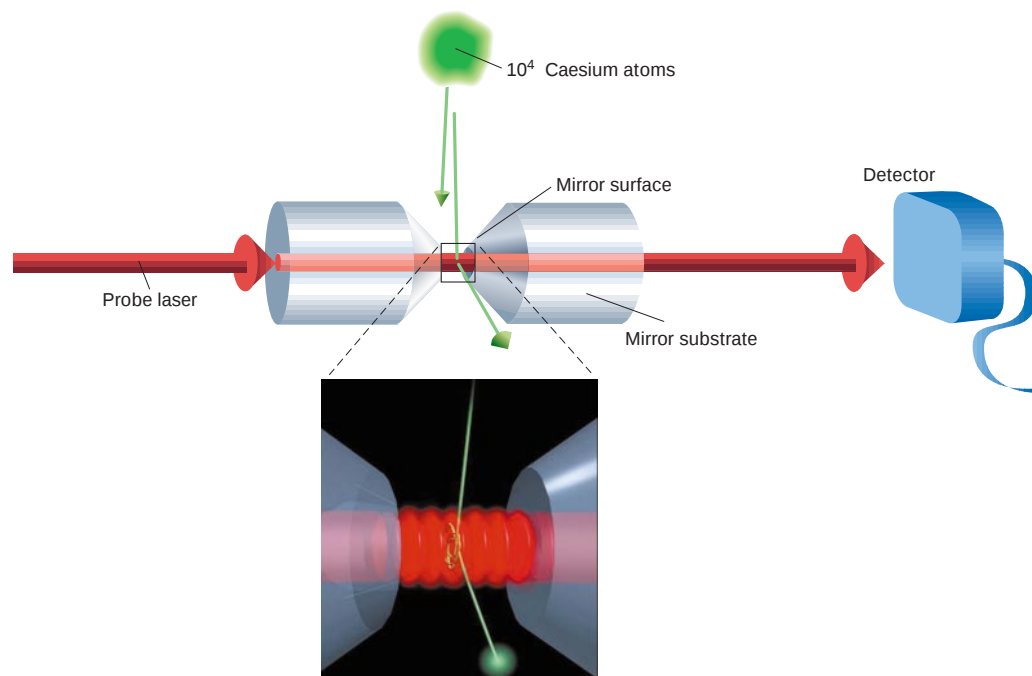
systems. Of particular interest is the reversible mapping of qubits from photon states to atomic states. This might allow the implementation of photonic quantum repeaters and also the teleportation of qubits from remotely located atoms through common EPR pairs of photons.

Cavity quantum electrodynamics and cold atoms

An elegant proposal to marry atomic and photonic quantum bits in optical cavity quantum electrodynamics (QED) was developed in 1997 (ref. 70). In this scheme, a small number of cold atoms are confined to the antinodes of a single-photon standing-wave field in a high-finesse Fabry–Perot optical cavity. Qubit states stored in the internal states of the atoms can be mapped to the qubit spanned by the number of photons (0 or 1) in the cavity through application of an appropriate laser pulse from the side. The photon can be made to leak out of the cavity within a pre-set time window, resulting in an ideal single-photon source for use in quantum communication⁷¹. Moreover, after the photon leaks out of the cavity, it can be deterministically 'caught' in a second cavity by the application of another laser pulse, whose envelope is time reversed with respect to the first pulse, as depicted in Fig. 6. Generalizations involving more than one atom in each cavity can distribute entanglement to many nodes while also featuring fault-tolerant error correction⁷².

Applying cavity-QED techniques to quantum information requires a single-atom/single-photon coupling that overwhelms any loss or decoherence rate, including atomic spontaneous emission and photon leakage through the mirrors. This leads to cavities of small

Figure 7 Single-atom cavity-QED experiment. Cold caesium atoms are dropped into a high-finesse optical cavity of axial spacing $\sim 10\ \mu\text{m}$. The trajectory of a single atom traversing the cavity is reconstructed (inset) by monitoring the field that leaks out of the cavity. (Image courtesy of H. J. Kimble, CalTech)



volume (for a high single-photon intensity) and extremely high mirror reflectivity (to keep the photon in the cavity). Groups from CalTech^{73,74} and MPQ⁷⁵ are beginning to master this difficult technology by trapping individually laser-cooled atoms inside miniature high-finesse cavities (Fig. 7). The atoms evolve in the requisite ‘strong-coupling’ limit, where even single photons in the cavity can significantly influence the dynamics of atomic motion⁷⁶. However, the atomic trajectories are still extended when compared to the optical wavelength, which effectively randomizes the atom–photon coupling and leads to decoherence. A second MPQ group is progressing on a complementary system, where individual atomic ions have been tightly bound within a larger optical cavity so that the atomic trajectory is well under an optical wavelength⁷⁷. But here the atom–photon system is not strongly coupled, because the insulating surfaces of a smaller cavity could interfere with the ion-trapping fields. These experiments, along with approaches from several other groups, point the way towards an ultimate goal in quantum information networking—that of coherently exchanging single atoms and single photons.

‘Stopping’ light

Several recent experiments from the universities of Harvard^{78,79} and Berkeley⁸⁰ have reported slowing and even ‘stopping’ light using the techniques of electromagnetically induced transparency⁸¹ in an atomic vapour. The stopping of light might seem trivial at first glance—it can be easily accomplished by putting an opaque black screen in front of a light source. Instead of irreversibly destroying the photons, however, these experiments offer the potential for mapping quantum states of photons onto collective quantum states of the atomic vapour⁸². For example, single-photon qubits can be mapped onto a single qubit spanning all N spins in a vapour:

$$[\alpha|0\rangle_{\gamma} + \beta|1\rangle_{\gamma}]|0_{\text{exc}}\rangle \rightarrow |0\rangle_{\gamma}[\alpha|0_{\text{exc}}\rangle + \beta|1_{\text{exc}}\rangle] \quad (4)$$

where $|1\rangle_{\gamma}$ and $|0\rangle_{\gamma}$ represent the presence or absence of the photon, respectively, and $|0_{\text{exc}}\rangle = |\downarrow_1\downarrow_2\downarrow_3\ldots\downarrow_N\rangle$ and $|1_{\text{exc}}\rangle = [|\uparrow_1\downarrow_2\downarrow_3\ldots\downarrow_N\rangle +$

$|\downarrow_1\uparrow_2\downarrow_3\ldots\downarrow_N\rangle + \ldots + |\downarrow_1\downarrow_2\downarrow_3\ldots\uparrow_N\rangle]/N^{1/2}$ are the symmetric states of zero and one excitation, respectively. Recent work from Harvard has demonstrated classical coherence in this process⁸³, although quantum coherence involving small numbers of photons has not yet been observed.

A drawback in this mapping is the requirement of quantum number states of photons in the first place, which may ultimately require the cold atom and cavity-QED systems mentioned above. One way around this requirement is to exploit interactions between the atoms in the vapour to generate quantum states of excitation. If a classical laser pulse excites atoms to Rydberg states for instance, then a strong resonant dipole–dipole interaction can inhibit (shift spectroscopically) further excitations, leaving the state $|1_{\text{exc}}\rangle$, or exactly one excited Rydberg atom in the collection, a phenomenon called the ‘dipole blockade’⁸⁴.

An alternative quantum network might map the continuous degrees of freedom of a light field (for example, polarization or electric-field quadratures) onto a collection of atoms. In this way, atoms can be entangled in symmetric (‘spin-squeezed’) states with many excitations, an extension of states like $|1_{\text{exc}}\rangle$ above^{85,86}. A group at the University of Aarhus has recently entangled spatially separated collections of atoms through common interaction with a light field that traverses both collections (ref. 87; and see Fig. 8). In this case, a macroscopic number of atoms is made to behave as a single quantum degree of freedom, and such a system has promise for teleporting states between remotely located quantum memories⁸⁸. These experiments can be remarkably simple, requiring only trivial optical equipment and thermal samples of atoms in vapour cells.

Conclusions

Quantum information technology is likely to have an important role in information processing after the demise of Moore’s Law. The extent of this role is unknown, as quantum information processors currently have only a limited number of known applications, such as number factoring, database searches and enhanced communication

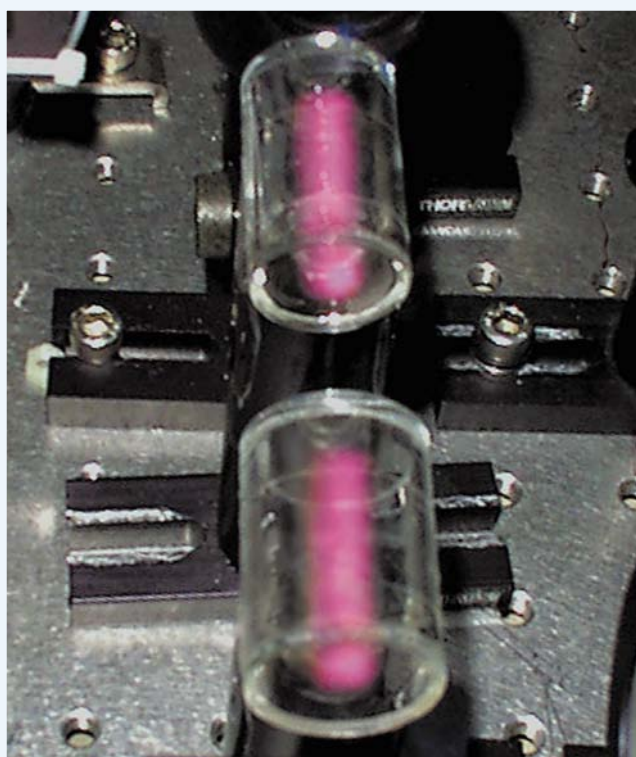


Figure 8 Pair of rubidium vapour cells used for storage and entanglement. A common laser beam traverses both cells, resulting in the purple fluorescence glow. Particular measurements of the polarization of the output light projects the atom collections in a 'spin-squeezed' state where the collective magnetization of the two macroscopic samples is weakly entangled. (Image courtesy of E. Polzik, University of Aarhus.)

protocols. Moreover, it may take decades to learn how to build large-scale quantum hardware, as this revolutionary form of information processing demands that revolutionary and exotic devices be engineered. Current devices come primarily from the areas of quantum optics and atomic physics, usually involving laser-cooled and trapped atoms. But perhaps the most exciting feature of this field is that the first large-scale quantum computer will probably be built from a physical system that is not currently known.

To quote Richard Feynman: "I think it is safe to say that nobody understands quantum mechanics." Current experiments that control individual atoms and photons will continue to lead the bizarre features of quantum-mechanical foundations to the forefront. After all, the systems currently under study are exactly the thought experiments envisioned by Einstein, Bohr and the other founding fathers of quantum physics. With the new language of quantum information, we might hope to gain more insight in the underlying quantum-physical principles, exactly as Shannon's theory of classical information ushered advances in physics responsible for the current digital age. □

- Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423; 623–656 (1948).
- Benioff, P. The computer as a physical system: a microscopic quantum mechanical model of computers as represented by Turing machines. *J. Stat. Phys.* **22**, 563–591 (1980).
- Benioff, P. Quantum mechanical Hamiltonian models of Turing machines that dissipate no energy. *Phys. Rev. Lett.* **48**, 1581–1585 (1982).
- Feynman, R. Simulating physics with computers. *J. Theor. Phys.* **21**, 467–488 (1982).
- Feynman, R. P. There's plenty of room at the bottom. *Eng. Sci.* February 1960 issue (produced by the CalTech Office of Public Relations) (1960); also available at <<http://www.zyvex.com/nanotech/feynman.html>>.
- Landauer, R. in *Proc. Drexel-4 Symp. Quantum Nonintegrability* (eds Feng, D. H. & Hu, B.-L.) 44–53 (International Press, Boston, 1997).
- Jozsa, R. in *The Geometric Universe* (ed. Huggett, S. et al.) 369–379 (Oxford Univ. Press, Oxford, 1998); preprint quant-ph/9707034 (Entanglement and quantum computation) at <http://xxx.lanl.gov> (1997).
- DiVincenzo, D. The physical implementation of quantum computation. *Fortschr. Phys.* **48**, 771–783 (2000).

- Special issue on experimental proposals for quantum computation. *Fortschr. Phys.* **48** (2000).
- DiVincenzo, D. Two-bit gates are universal for quantum computation. *Phys. Rev. A* **51**, 1015–1022 (1995).
- Lloyd, S. Almost any quantum logic gate is universal. *Phys. Rev. Lett.* **75**, 346–349 (1995).
- Einstein, A., Podolsky, B. & Rosen, N. Can quantum-mechanical description of reality be considered complete? *Phys. Rev.* **47**, 777–780 (1935).
- Kahn, D. *The Codebreakers: The Story of Secret Writing* (Macmillan, New York, 1967).
- Bennett, C. & Brassard, G. Quantum cryptography: public key distribution and coin tossing. *Proc. IEEE Int. Conf. Comp. Syst. Signal Proc.* **11**, 175–179 (1984).
- Wiesner, S. Conjugate coding. *Sigact News* **15**, 78–88 (1983).
- Bennett, C. Quantum cryptography using any two nonorthogonal states. *Phys. Rev. Lett.* **68**, 3121–3124 (1992).
- Franson, J. & Ilves, H. Quantum cryptography using optical fibers. *Appl. Opt.* **33**, 2949–2954 (1994).
- Marand, C. & Townsend, P. Quantum key distribution over distances as long as 30 km. *Opt. Lett.* **20**, 1695–1697 (1995).
- Müller, A., Zbinden, H. & Gisin, N. Quantum cryptography over 23 km in installed under-lake telecom fiber. *Europhys. Lett.* **33**, 335–339 (1996).
- Hughes, R. J., Morgan, G. L. & Peterson, C. G. Quantum key distribution over a 48 km optical fibre network. *J. Mod. Opt.* **47**, 533–547 (2000).
- Ekert, A. K. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.* **67**, 661–663 (1991).
- Jennewein, T., Simon, C., Weihs, G., Weinfurter, H. & Zeilinger, A. Quantum cryptography with entangled photons. *Phys. Rev. Lett.* **84**, 4729–4732 (2000).
- Naik, D. S., Peterson, C. G., White, A. G., Berglund, A. J. & Kwiat, P. G. Entangled state quantum cryptography: eavesdropping on the Ekert protocol. *Phys. Rev. Lett.* **84**, 4733–4736 (2000).
- Tittel, W., Brendel, J., Zbinden, H. & Gisin, N. Quantum cryptography using entangled photons in energy-time Bell states. *Phys. Rev. Lett.* **84**, 4737–4740 (2000).
- Kwiat, P. G., Waks, E., White, A. G., Appelbaum, I. & Eberhard, P. H. Ultrabright source of polarization-entangled photons. *Phys. Rev. A* **60**, 773–776 (1999).
- Bennett, C. H. et al. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
- Braunstein, S. L. A fun talk on teleportation. <www.informatics.bangor.ac.uk/~schmuel/tport.html> (5 February 1995).
- Bouwmeester, D. et al. Experimental quantum teleportation. *Nature* **390**, 575–579 (1997).
- Boschi, D., Branca, S., De Martini, F., Hardy, L. & Popescu, S. Experimental realization of teleporting an unknown pure quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **80**, 1121–1125 (1998).
- Braunstein, S. L. & Kimble, H. J. A posteriori teleportation. *Nature* **394**, 840–841 (1998).
- Bouwmeester, D. et al. Reply to 'A posteriori teleportation' by S. L. Braunstein and H. J. Kimble. *Nature* **394**, 841 (1998).
- Furusawa, A. et al. Unconditional quantum teleportation. *Science* **282**, 706–709 (1998).
- Schrödinger, E. The present situation in quantum mechanics. *Naturwissenschaften* **23**, 807–812; 823–828; 844–849 (1935).
- Deutsch, D. Quantum theory, the Church-Turing principle and the universal quantum computer. *Proc. R. Soc. Lond. A* **400**, 97–117 (1985).
- Shor, P. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comp.* **26**, 1484–1509 (1997).
- Rivest, R., Shamir, A. & Adleman, L. A method for obtaining digital signatures and public-key cryptosystems. *Commun. Assoc. Comput. Mach.* **21**, 120–126 (1978).
- Grover, L. Quantum mechanics helps in searching for a needle in a haystack. *Phys. Rev. Lett.* **79**, 325–328 (1997).
- Shor, P. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A* **52**, 2493–2496 (1995).
- Steane, A. Error correcting codes in quantum theory. *Phys. Rev. Lett.* **77**, 793–797 (1996).
- Preskill, J. Battling decoherence: the fault-tolerant quantum computer. *Phys. Today* **52**, 24–30 (1999).
- Cirac, I. & Zoller, P. Quantum computations with cold trapped ions. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
- Steane, A. The ion trap quantum information processor. *Appl. Phys. B* **64**, 623–642 (1997).
- Wineland, D. et al. Experimental issues in coherent quantum manipulation of trapped atomic ions. *J. Res. Natl. Inst. Stand. Tech.* **103**, 259–328 (1998).
- Sørensen, A. & Mølmer, K. Quantum computation with ions in thermal motion. *Phys. Rev. Lett.* **82**, 1971–1975 (1999).
- Blatt, R. & Zoller, P. Quantum jumps in atomic systems. *Eur. J. Phys.* **9**, 250 (1988).
- Monroe, C., Meekhof, D. M., King, B. E., Itano, W. M. & Wineland, D. J. Demonstration of a universal quantum logic gate. *Phys. Rev. Lett.* **75**, 4714–4717 (1995).
- Sackett, C. et al. Experimental entanglement of four particles. *Nature* **404**, 256–259 (2000).
- Turchette, Q. A. et al. Heating of trapped ions from the quantum ground state. *Phys. Rev. A* **61**, 063418-1–063418-8 (2000).
- Larson, D. J. et al. Sympathetic cooling of trapped ions: a laser-cooled two-species nonneutral ion plasma. *Phys. Rev. Lett.* **57**, 70 (1986).
- Kielinski, D. et al. Sympathetic cooling of trapped ions for quantum logic. *Phys. Rev. A* **61**, 032310-1–032310-8 (2000).
- Morigi, G. & Walther, H. Two-species Coulomb chains for quantum information. *Eur. Phys. J. D* **13**, 261–269 (2001).
- Jessen, P. & Deutsch, I. Optical lattices. *Adv. At. Mol. Opt. Phys.* **37**, 95–138 (1996).
- Hamann, S. et al. Resolved-sideband Raman cooling to the ground state of an optical lattice. *Phys. Rev. Lett.* **80**, 4149–4152 (1998).
- Vuletic, V., Chin, C., Kerman, A. J. & Chu, S. Raman sideband cooling of trapped cesium atoms at very high densities. *Phys. Rev. Lett.* **81**, 5768–5771 (1998).
- Perrin, H., Kuhn, A., Bouchoule, I. & Salomon, C. Sideband cooling of neutral atoms in a far-detuned optical lattice. *Europhys. Lett.* **42**, 395–400 (1998).
- Raithel, G., Phillips, W. D. & Rolston, S. L. Coherence decay of wave-packets in optical lattices. *Phys. Rev. Lett.* **81**, 3615 (1998).
- Haycock, D. L., Alsing, P. M., Grondalski, J., Deutsch, I. H. & Jessen, P. S. Mesoscopic quantum coherence in an optical lattice. *Phys. Rev. Lett.* **85**, 3365 (2000).
- Brennen, G., Caves, C., Jessen, P. & Deutsch, I. Quantum logic gates in optical lattices. *Phys. Rev. Lett.* **82**, 1060–1063 (1999).

59. Jaksch, D., Briegel, H. J., Cirac, J. I., Gardiner, C. W. & Zoller, P. Entanglement of atoms via cold controlled collisions. *Phys. Rev. Lett.* **82**, 1975–1978 (1999).
60. Jaksch, D. *et al.* Fast quantum gates for neutral atoms. *Phys. Rev. Lett.* **85**, 2208–2211 (2000).
61. DeMille, D. Quantum computation with trapped polar molecules. Preprint quant-ph/0109083 at <http://xxx.lanl.gov> (2001).
62. Jaksch, D., Bruder, C., Cirac, J. I., Gardiner, C. W. & Zoller, P. Cold bosonic atoms in optical lattices. *Phys. Rev. Lett.* **81**, 3108–3111 (1998).
63. Spekkens, R. W. & Sipe, J. E. Spatial fragmentation of a Bose-Einstein condensate in a double-well potential. *Phys. Rev. A* **59**, 3868–3877 (1999).
64. Orzel, C., Tuchman, A. K., Fenselau, M. L., Yasuda, M. & Kasevich, M. A. Squeezed states in a Bose-Einstein condensate. *Science* **291**, 2386–2389 (2001).
65. Greiner, M., Mandel, O., Esslinger, T., Hansch, T. W. & Bloch, I. Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms. *Nature* **415**, 39–44 (2002).
66. Cirac, I. & Zoller, P. A scalable quantum computer with ions in an array of microtraps. *Nature* **404**, 579–581 (2000).
67. Monroe, C., Meekhof, D. M., King, B. E. & Wineland, D. J. A Schrodinger cat superposition state of an atom. *Science* **272**, 1131–1136 (1996).
68. Bennett, C. H. *et al.* Purification of noisy entanglement and faithful teleportation via noisy channels. *Phys. Rev. Lett.* **76**, 722–725 (1996).
69. Briegel, H.-J., Dur, W., Cirac, J. I. & Zoller, P. Quantum repeaters: the role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
70. Cirac, I., Zoller, P., Kimble, H. & Mabuchi, H. Quantum state transfer and entanglement distribution among distant nodes in a quantum network. *Phys. Rev. Lett.* **78**, 3221–3224 (1997).
71. Law, C. & Kimble, H. Deterministic generation of a bit-stream of single-photon pulses. *J. Mod. Opt.* **44**, 2067–2074 (1997).
72. van Enk, S., Cirac, I. & Zoller, P. Ideal quantum communication over noisy channels, a quantum optical implementation. *Phys. Rev. Lett.* **78**, 4293–4296 (1997).
73. Ye, J., Vernooy, D. W. & Kimble, H. J. Trapping of single atoms in cavity QED. *Phys. Rev. Lett.* **83**, 4987–4990 (1999).
74. Hood, C. J., Lynn, T. W., Doherty, A. C., Parkins, A. S. & Kimble, H. J. The atom-cavity microscope: single atoms bound in orbit by single photons. *Science* **287**, 1447–1453 (2000).
75. Pinske, P., Fischer, T., Maunz, P. & Rempe, G. Trapping an atom with single photons. *Nature* **404**, 365–368 (2000).
76. Doherty, A., Lynn, T. W., Hood, C. J. & Kimble, H. J. Trapping of single atoms with single photons in cavity QED. *Phys. Rev. A* **63**, 013401–013401-24 (2001).
77. Guthohrlein, G. R., Keller, M., Hayasaka, K., Lange, W. & Walther, H. A single ion as a nanoscopic probe of an optical field. *Nature* **414**, 49–51 (2001).
78. Liu, C., Dutton, Z., Behroozi, C. H. & Hau, L. V. Observation of coherent optical information storage in an atomic medium using halted light pulses. *Nature* **409**, 490–493 (2001).
79. Phillips, D. F., Fleischhauer, A., Mair, A., Walsworth, R. L. & Lukin, M. D. Storage of light in atomic vapor. *Phys. Rev. Lett.* **86**, 783–786 (2001).
80. Budker, D., Kimball, D. F., Rochester, S. M. & Yashchuk, V. V. Nonlinear magneto-optics and reduced group velocity of light in atomic vapor with slow ground state relaxation. *Phys. Rev. Lett.* **83**, 1767–1770 (1999).
81. Lukin, M. & Imamoglu, A. Controlling photons using electromagnetically induced transparency. *Nature* **413**, 273–276 (2001).
82. Fleischhauer, M. & Lukin, M. Dark-state polaritons in electromagnetically induced transparency. *Phys. Rev. Lett.* **84**, 5094–5097 (2000).
83. Mair, A., Hager, J., Phillips, D. F., Walsworth, R. L. & Lukin, M. D. Phase coherence and control of stored photonic information. Preprint quant-ph/0108046 at <http://xxx.lanl.gov> (2001).
84. Lukin, M. *et al.* Dipole blockade and quantum information processing in mesoscopic atomic ensembles. *Phys. Rev. Lett.* **87**, 037901–1–037901-4 (2001).
85. Hald, J., Sørensen, J. L., Schori, C. & Polzik, E. S. Spin squeezed atoms: a macroscopic entangled ensemble created by light. *Phys. Rev. Lett.* **83**, 1319–1322 (1999).
86. Kuzmich, A., Mandel, L. & Bigelow, N. Generation of spin squeezing via continuous quantum nondemolition measurement. *Phys. Rev. Lett.* **85**, 1594–1597 (2000).
87. Julsgaard, B., Kozhekin, A. & Polzik, E. Experimental long-lived entanglement of two macroscopic objects. *Nature* **413**, 400–403 (2001).
88. Duan, L.-M., Lukin, M. D., Cirac, J. I. & Zoller, P. Long-distance quantum communication with atomic ensembles and linear optics. *Nature* **414**, 413–418 (2001).
89. Wolf, F. A. *Taking the Quantum Leap* (Harper and Row, San Francisco, 1981).
90. Zurek, W. H. Decoherence and the transition from quantum to classical. *Phys. Today* **44**, 34–44 (1991).
91. Braunstein, S. L., D'Ariano, G. M., Milburn, G. J. & Sacchi, M. F. Universal teleportation with a twist. *Phys. Rev. Lett.* **84**, 3486–3489 (2000).

Acknowledgements

This work is supported by the US National Security Agency and Advanced Research and Development Activity under an Army Research Office contract, and by the National Science Foundation ITR programme.

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Group European Manager:

Leonie Welss (4954)

European Manager:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),
Andy Douglas (4975)

Holland/ Italy:

Nevin Bayoumi (4978)

Scandinavia:

Sille Opstrup (4994)

Spain/ Portugal:

Leonie Welss (4954)

Production Manager:

Billie Franklin

To send materials use London

address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau
Tel + 33 (0) 1 43 20 16 51
Fax + 33 (0) 1 43 20 51 52
e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mails: p.phelan@nature.com
k.turner@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Director:

Ben Crowe

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Ashly de Leon

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harakatomachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
e-mail: k.cowan@naturejp.com
Japan Manager: Kate Cowan

naturejobs

Translating words into action

Language can be either a barrier or a bridge in recruitment. In German-speaking countries, there are examples of it being both. A recent report from the European Commission (EC) shows how language can be an obstacle. The report examines where postgraduates and postdoctoral fellows end up under the EC's Marie Curie Fellowships, which enable young scientists to study in European countries other than their own. Only 10% of the 2,080 fellowships awarded between 1999 and 2001 were used to study at a German university, research institute or industry research programme. In response, German researchers say that language difficulties are partly to blame (see *Nature* 415, 945; 2002).

But in recent visits to Germany and Austria, I saw how operating labs under a common language — English — can have the opposite effect. Visits to the European Molecular Biology Laboratory in Heidelberg and the Research Institute of Molecular Pathology in Vienna were, to me, telling.

Both labs employ an international mix of young scientists, some of whom said that the English-based science culture was instrumental in attracting them to their current posts. The need to learn German from scratch in order to work at either of the labs would have been an impediment, they said, although they were interested in learning the language to help them with everyday life outside the lab.

Choosing English as a scientific language can sometimes be controversial. Perhaps that's because cultural identity and linguistic identity can get completely confused. Although it may be better to tackle Goethe in German, it might be more practical to discuss the intricacies of cell signalling or gene expression in English. And opting for English as a common language may help to attract a more international mix to the lab.

Paul Smaglik

Naturejobs editor



Contents

CAREERS AND RECRUITMENT

Condensed-matter physics
faces up to recession p4

Japan's brave new world p6

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

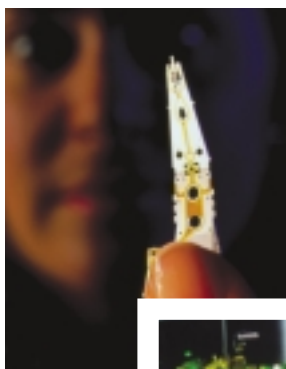
FOCUS

SPOTLIGHT

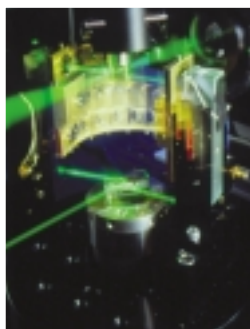
RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



Applications matter



Basic research in condensed-matter physics continues to lose ground to focused R&D. Effective cross-disciplinary partnering between universities, government labs and industry is seen as essential to the future health of basic research in this field. Brendan Horton explores.

IBM

Most companies have found long-term basic research in condensed-matter and material physics to be largely unsustainable if they are to remain globally competitive. Gone are the days when US companies such as General Electric, Ford and Xerox did fundamental research. Today only a handful of industrial laboratories are still in the game — even though the distance between basic research and technology development is relatively short.

Despite the high quality of basic research that still goes on at Lucent Technologies' Bell Laboratories and IBM Research, only a small portion ever adds to the bottom line on their balance sheet. Many patents do not realize much — if any — revenue, and many discoveries, such as optical tweezers or the scanning tunnelling microscope, become research tools rather than providing significant financial returns.

So as industry continues to sharpen its focus on short-term product development, the job prospects outside academia for physicists in this field are increasingly found in research and development (R&D) rather than basic research. To make matters worse, US federal investment in this area declined by more than 10% between 1985 and 1997 (excluding government expenses for stewardship of the national laboratories), according to a report by the US National Research Council's Committee on Condensed-Matter and Materials Physics.

INDUSTRIAL EXPERIENCE

In spite of funding problems for basic research in both industry and academia, the United States remains attractive to young scientists from Europe looking for postdoctoral experience or summer internships in industry, says Hans Coufal, manager of science and technology at IBM Research's Almaden Research Center in San Jose, California. Coufal himself originally became involved with IBM Research as a

Start-up gets a jump-start

Some start-ups, such as Kodeos Communications of South Plainfield, New Jersey, formed in January 2001 by former Bell Labs researchers Jason Stark and Gadi Lenz, seem to be weathering the present economic slump. In fact, last August the company raised US\$12 million in equity financing.

Stark, a former Bell Labs researcher, says that — in contrast to the early-to-mid 1990s — the current economic downturn has led to a general reduction in the number of research positions in condensed-

matter physics, optical physics and materials science. The booming market of the 1990s attracted a lot of highly skilled researchers to the optical communications industry, he says. But subsequent corporate downsizing has put many out of work.

For now, conserving capital is a priority for Kodeos. The company is not actively pursuing new staff. But it is always on the lookout for individuals with expertise in high-speed digital and analog design and in signal processing.

B.H.

summer intern, visiting from Germany, where he attended the Technical University of Munich. This experience convinced him to return to IBM as a visiting scientist and, in 1981, he joined as a research staff member, pursuing research in laser-induced, sonar-acoustic transients.

"It is still typical for someone pursuing an academic career in Europe to spend a year in a foreign country, usually as a postdoc," says Coufal, who recommends the United States. "In the United States, you get responsibility much earlier in your career than you do in a country like Germany, where seniority is very important," he says. Once in the United States, it is easy to get hooked by the way science is done, the lifestyle, or even the climate, he adds.

KODEOS



Jason Stark feels physics has been affected by the economic downturn.

CAREERS AND RECRUITMENT

The United Kingdom, too, has had a hard time holding on to its recent graduates, with far fewer opportunities for postdoctoral work or summer internships in industry than in, for example, the United States or continental Europe. With the exception of 'soft' condensed-matter physics, where there are research jobs in complex fluids, macromolecular and biological systems tailored to the pharmaceutical and chemical industries, Britain has very little industry in which to employ condensed-matter physicists, says Peter Littlewood of the Cavendish Laboratory at the University of Cambridge.

"In contrast to the United States, Britain has an all-or-nothing structure: either become a professor or go into consultancy and banking. The big condensed-matter employers in the United States and other parts of Europe — semiconductors and telecoms — don't exist here," says Littlewood. It is not just that there is a shortage of industry jobs for postdocs, he says. He has also noticed that those who are academically inclined are going to the United States. Two years ago he advertised for a postdoc and was stunned to see that, out of numerous applicants, only two were British.

REVERSING THE FLOW

In 1998, the US National Science Foundation (NSF) reported that, in the United States, 29% of the science-and-engineering workforce holding doctoral degrees were foreign nationals. This once represented a one-way flow of intellectual talent, but, in its study, the NSF found that roughly half of all foreign graduate students returned home after receiving their PhD, and more returned home after some time teaching or training.

In Canada, the National Research Council (NRC) has been taking action to retain and reclaim some of its brightest intellects, as well as recruiting foreigners as needed. One of its latest initiatives was announced late last year with the creation of the National Institute of Nanotechnology (NINT), an unprecedented partnership between the Canadian federal government, the province of Alberta and the University of Alberta (see *Nature* **412**, 846; 2001).

NINT is expected to be one node in a network of national R&D facilities in nanotechnology. The NRC

and Alberta's provincial government will each put up half of the Can\$120 million (US\$75 million) start-up costs for the new institute, which will open in July in a temporary space. The institute's new facilities, slated for completion in about 2005, will be built on the campus of the University of Alberta in Edmonton. When fully operational, it will house about 40 researchers with a total staff of around 150 people, says NINT's director Dan Wayner.

It will host scientists from a range of disciplines, from chemical, mechanical and electrical engineering through to the biological and medical sciences. Researchers at the institute will hold joint appointments with the university, and vice versa. It is also anticipated that partnering with universities will bring some 200 students to work at the institute each year.

"The idea is to build on our strengths and to align our research programmes so that we don't unnecessarily duplicate expertise and facilities," says Wayner, formerly of the NRC's Steacie Institute for Molecular Sciences (SIMS) in Ottawa.

Wayner is no stranger to taking a multidisciplinary approach to research. With a background in organic and physical organic chemistry, he has for some time enjoyed a fruitful collaboration with SIMS colleague Robert Wolkow, who has expertise in both physics and surface science (see IBM, Bell and home again, below).

There will be an 'incubator' facility for fledgling companies at the institute, says Wayner. "I think it's important that we nurture those small enterprises as they start to spin out." He also notes the importance of sharing any equity that accrues from these companies. Wayner anticipates that the institute will stimulate economic growth in industries centred around nanotechnology on a 10- to 15-year timeframe.

Brendan Horton is a freelance writer based in Maryland.

Web links

Steacie Institute for Molecular Sciences ♦ www.sims.nrc.ca/sims

National Institute for Nanotechnology ♦ www.nrc.ca/nanotech



Hans Coufal: the US is still attractive to physicists.



Britain cannot currently compete in the condensed-matter marketplace, says Peter Littlewood (right).

IBM, Bell and home again

Robert Wolkow's research career has included stints at both IBM Research and Bell Labs. But it was the changing research climate at Bell Labs and the offer of a job as head of the scanning tunnelling microscopy (STM) group at the National Research Council's Steacie Institute for Molecular Sciences (SIMS) in Ottawa that enticed him back to his native Canada in 1992.

Wolkow began postdoc work at IBM Research in Yorktown Heights, New York.

He switched to Bell after giving a seminar on STM at the company. He encountered a highly charged, somewhat outrageous, atmosphere. "People would just holler out questions and take shots at people, but I liked that and I'd fire right back," he says.

Although Wolkow was the quiet Canadian, he found the research environment at Bell Labs stimulating. But by 1991, the business units within AT&T were expressing their dissatisfaction with Bell Labs.

Disillusioned with the erosion of what he saw as an ideal research environment, Wolkow accepted the position at SIMS, after nearly deciding to get out of research altogether.

Wolkow might never have returned to Canada were it not for substantial federal investment by the National Research Council in creating a high-quality government institute aimed at recruiting and keeping its top scientists. At SIMS, this investment is geared towards long-term

interdisciplinary research in areas of molecular science, with the broader goal of stimulating Canada's economy.

Since leaving Bell Labs, Wolkow says he has thought a lot about possible models for doing research. "Although I often talk about Bell Labs like it was some kind of ideal, I realize that it is not something that could be recreated from scratch," he says. "The nearest you can do is to pay substantial salaries, I think, and provide excellent freedom and resources." B.H.

Slim pickings for silicon specialists

Electronics companies in Japan are slashing jobs for silicon physicists — once the veritable élite of the country's corporate R&D world. And public-sector research is only slowly picking up, says Robert Triendl.

The employment prospects for condensed-matter physicists in Japan today are less promising than they used to be only a few years ago, despite much attention to emerging research fields such as nanotechnology and quantum computing. Especially in the field of silicon technology, corporate downsizing has forced many researchers to change career or to try the often-difficult transition to academia or public-sector research.

Despite much recent interest in new nanomaterials, such as carbon nanotubes, there is little doubt that silicon will remain the material of choice in the semiconductor industry for many years to come. But being a silicon specialist these days in Japan is no longer a guarantee of a research career in industry — far from it.

During the 1980s, when Japanese electronics companies were reaping ever-larger shares of the global market for semiconductor devices and especially memory chips, many companies invested heavily in fundamental research on future generations of semiconductor devices. But, faced with increasing competition from cheap suppliers in Korea or Taiwan and a continuing economic downturn, many electronics and telecommunications companies in Japan started to reorganize their research and development (R&D) activities during the mid-1990s, shifting away from hardware manufacturing towards software and services.

Meanwhile, companies in other industries that had entered the semiconductor field during the 1980s — including many of Japan's steel companies — abandoned their activities, selling off factories and closing down most of their research in condensed-matter physics.

Nippon Telegraph and Telephone (NTT), the former monopoly telecommunications carrier, first downsized its hardware-related activities during the company's reorganization in 1999. Hardware-related activities at NTT Research Laboratories, once a major centre for research in silicon technology in

Japan that operated its own synchrotron radiation facility, were severely cut in favour of R&D in software and systems.

Today, R&D spending at NTT is down from more than ¥300 billion (US\$2.2 billion) at its peak in the early 1990s to less than ¥200 billion. Many research scientists in silicon technology have left the company to work in academia or else have transferred into new business areas, such as the Internet or international services. All remaining research into materials and devices has been transferred to a small basic-research laboratory that is focused on 'far out' areas of research, such as quantum computing.

INDEPENDENT ATTITUDE

Over the past few years, the number of staff at NTT Basic Research Laboratories has remained constant, points out the labs' director Sunao Ishihara, who used to work on the development of new lithographic techniques for semiconductor manufacturing, such as X-ray lithography. But the 125 staff represent the remnants of a once burgeoning hardware-research activity that, at its height in the late 1980s, employed hundreds of silicon specialists.

Despite a continuing commitment by the NTT group to basic research, maintaining a continuous stream of research funding for this kind of research is no easy task, says Ishihara. Today, NTT's basic-research laboratories, like the rest of NTT research, are an independent company under the NTT holding company. But, whereas other group companies — such as NTT DoCoMo, the highly successful mobile carrier — have their own research facilities, the troubled local NTT companies have little need for advanced R&D, let alone basic research.

NTT is not the only company that has restructured its R&D activities over the past few years. Many of the basic research centres set up by Japanese electronics companies during the 1980s, when money was readily available, have gone through successive waves of



Sunao Ishihara: it is hard to maintain funding.

CAREERS AND RECRUITMENT

restructuring and downsizing. Some have disappeared altogether. In Japan, the amount of government funding for basic research in industry is tiny and usually comes in the form of large, highly inflexible cooperation projects. For corporate basic science labs, this leaves very little room for manoeuvre if economic conditions worsen.

Hitachi's Advanced Research Laboratory, a basic-research laboratory set up in the late 1980s, is now largely deserted after the remaining scientists in its biology division were transferred to a new company — Hitachi Life Science Group — some two years ago. At its peak, the laboratory had employed some 250 researchers, about half of them in condensed-matter physics or instrumentation. Even today, the lab's facilities are among the best available and include the world's most powerful electron microscope.

The outlook suggests there will be more difficult times ahead. Over the past few months, Japanese

electronics firms have come under pressure once again to restructure after revenues dropped drastically as a result of the downturn in the information technology (IT) markets and the burst of the US Internet bubble. All major electronics firms in Japan have announced plans to accelerate the shift from manufacturing to services — and to retrain thousands of hardware engineers and researchers as systems engineers.

HARD CHOICES

Whether such in-house training programmes can produce the kind of high-grade IT professionals that will enable electronics companies to land large outsourcing contracts, and increase their revenues from services, is a question that is hotly debated among industry analysts. For research scientists, the choice is increasingly between leaving the company or going through several months of retraining and work as a systems engineer or security specialist.

A small number of industry scientists with a track-record in emerging research areas where government funding is increasing — such as nanotechnology or quantum computing — have found their way back to employment in academia or public-sector research. Nevertheless, the academic job market in Japan remains tight. Especially in the field of silicon technology, one of the areas hardest hit by corporate downsizing, career opportunities within universities have traditionally been limited and are confined to a few laboratories and centres, such as the Research Center for Nanodevices and Systems at Hiroshima University.

Scientists with industry experience could play a crucial role in linking industry activities and academia, says Masataka Hirose, who founded the Research Center for Nanodevices and Systems. Hirose now heads the Millennium Research for Advanced Information Technology project, or Mirai ('future'), a new government-funded research effort aiming to develop future generation semiconductor technologies (see 'Silicon futures', left).

So far, incentives for Japanese university staff to collaborate more closely with industry have been limited. But, with the reorganization of public universities into semi-private agencies, universities in Japan are coming under increasing pressure to engage in research activities more directly linked to economic and social return. Although connections and support from powerful university professors — rather than a publications track record or skill as an experimentalist — still often determine who will be offered a position, many expect that the new status for universities will lead eventually to a more competitive market for academic careers. ■

Robert Triendl is a freelance writer based in Tokyo.

Web links

NTT Basic Research Laboratories

◆ www.brl.ntt.co.jp

Hitachi Advanced Research Laboratory

◆ www.harl.hitachi.co.jp

Advanced Semiconductor Research Center

◆ unit.aist.go.jp/asrc/asrc-e



Masataka Hirose wants to see more flexibility.

Silicon futures

Once completed later this year, some 250 scientists will work in a new clean-room facility located just behind the Tsukuba campus of the National Institute for Advanced Industrial Science and Technology. The facility will host the government-financed Mirai project and its industry-financed companion, Asuka — initiatives aimed at regaining Japan's competitive edge in advanced semiconductor manufacturing technologies.

Researchers at Mirai are working on technologies that are still many years from commercialization, whereas at Asuka the focus is on those areas that are near production. Researchers at Asuka usually come from member companies and work at the joint facility for a limited period of time. Mirai has also hired a number of industry researchers, as well as academic scientists with industry experience, says the project's director Masataka Hirose.

According to Hirose, to be successful the new

facility must combine a strong goal orientation with flexibility. Rigid management and a lack of flexibility, together with inadequate evaluation and insufficient attention to intellectual property issues, have been among the problems that have hampered Japanese collaborative research efforts in the past — very few of which have been successful. As Japan makes switches from a manufacturing economy towards a knowledge-based economy, new organizational approaches are crucial, Hirose says.

At Asuka, the diversification of semiconductor activities among its member companies has already put some of the project's objectives and its organizational principles in question. Faced with increasingly diverse interests on the part of its member companies, some argue that a shift towards an organization that emphasizes contracts with individual companies — rather than joint R&D — is now inevitable. R.T.

Well-equipped: Hitachi's laboratories feature the world's most powerful electron microscope.

